BUSINESS INTELLIGENCE

MBA II-Year IV-Sem (Osmania University)

SIA GROUP
ENLIGHTENING LIVES
THROUGH EDUCATION
www.universalbooks.com

CONTENTS

Unit- **1**

ESSAY QUESTIONS

Q1. What is Business Intelligence? Discuss in detail about its history and evolution.

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.2, Q.No. 1.

Q2. List the styles of business intelligence and also discuss its benefits.

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.3, Q.No. 2.

Q3. Discuss in detail about Automated Decision systems (ADS).

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.4, Q.No. 3.

Q4. Discuss in detail about real-time business intelligence.

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.4, Q.No. 4.

Q5. Write about business intelligence value chain.

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.6, Q.No. 5.

Q6. Explain the architecture of business intelligence.

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.7, Q.No. 6.

SHORT QUESTIONS

Q1. Write a short note on business intelligence.

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.9, Q.No. 1.

Q2. List the benefits of BI.

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.9, Q.No. 3.

Q3. What are the tools and techniques involved in business analytics?

Answer :   *Important Question*

For answer refer Unit-I, Page No. 1.10, Q.No. 7.

Unit-2

ESSAY QUESTIONS

Q1.   What is a data warehouse? Discuss its characteristics.

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.2, Q.No. 1.

Q2.   Describe various types of data warehouses.

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.3, Q.No. 2.

Q3.   Discuss in brief about data integration.

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.9, Q.No. 8.

Q4.   Give a simple taxonomy of data in data mining.

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.14, Q.No. 14.

Q5.   Explain various classification techniques of data mining.

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.16, Q.No. 17.

Q6.   List out various software tools of data mining.

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.20, Q.No. 20.

SHORT QUESTIONS

Q1.   Define data warehouse.

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.25, Q.No. 1.

Q2.   What are the advantages of data warehouse?

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.25, Q.No. 3.

Q3.   What is meant by classification? What are applications of classification model?

Answer :   *Important Question*

For answer refer Unit-II, Page No. 2.26, Q.No. 10.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD. ——————————

Unit-3

ESSAY QUESTIONS

Q1. Define Business Performance Management (BPM)? State the features and components of BPM.

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.2, Q.No. 1.

Q2. Explain in detail about BPM cycle.

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.3, Q.No. 3.

Q3. What is performance measurement system? What are the characteristics of an effective performance measurement system?

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.7, Q.No. 8.

Q4. What are the two widely used approaches that support the basic processes underlining BPM?

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.9, Q.No. 11.

Q5. What is logical system architecture? What are the basic parts/elements of a BPM architecture?

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.15, Q.No. 16.

Q6. What do you mean by dashboards? Differentiate between dashboards and scorecards. What layers of information are provided by the dashboards?

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.17, Q.No. 19.

SHORT QUESTIONS

Q1. Write briefly about TSA's PIMS?

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.19, Q.No. 1.

Q2. State the key elements of a diagnostic control system?

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.19, Q.No. 3.

Q3. Distinguish between lean production and six sigma.

Answer :   *Important Question*

For answer refer Unit-III, Page No. 3.20, Q.No. 5.

ESSAY QUESTIONS

Q1. Define analytics and Business Analytics (BA). State the prerequisites for effective business analytics.

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.2, Q.No. 1.

Q2. Discuss in detail tools and techniques of Business Analytics (BA).

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.3, Q.No. 2.

Q3. What is the usage of Business Analytics? What are the benefits of implementing Business Analytics in an Organization.

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.6, Q.No. 9.

Q4. Write in short about data visualization and visualization spreadsheets.

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.7, Q.No. 11.

Q5. Explain about GIS.

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.8, Q.No. 13.

Q6. Write about GIS Vs GPS.

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.9, Q.No. 14.

SHORT QUESTIONS

Q1. Explain briefly why Business Intelligence/Business Analytics projects fail.

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.11, Q.No. 2.

Q2. Define Data visualization.

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.11, Q.No. 3.

Q3. What is GPS?

Answer :   *Important Question*

For answer refer Unit-IV, Page No. 4.12, Q.No. 6.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

ESSAY QUESTIONS

Q1. What are the factors that affect the implementation of business intelligence. Also, list its critical success factors.

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.2, Q.No. 1.

Q2. Discuss in detail about the managerial issues related to BI implementation.

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.3, Q.No. 2.

Q3. Discuss in detail about Web 2.0 as an emerging trends in Business Intelligence.

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.5, Q.No. 6.

Q4. Write about representative areas and examples of enterprise social networking.

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.9, Q.No. 11.

Q5. Discuss in detail about RFID.

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.12, Q.No. 14.

Q6. Discuss in detail about reality mining.

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.13, Q.No. 15.

SHORT QUESTIONS

Q1. Write short notes on,

    (i)   Functional integration

    (ii) Physical Integration.

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.15, Q.No. 1.

Q2. Discuss the benefits of collaborative decision making (CDM).

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.16, Q.No. 5.

Q3. What is Reality Mining?

Answer :   *Important Question*

For answer refer Unit-V, Page No. 5.16, Q.No. 8.

UNIT

1

Introduction to Business Intelligence (BI)

After studying this unit, one would be able to understand,

❖ The concept of Business Intelligence (BI).

❖ History and evolution of BI.

❖ Various styles and benefits of BI.

❖ The concept of Automated Decision Systems (ADS).

❖ The concept of Real-time BI.

❖ Business Intelligence value chain.

❖ The architecture of BI.

INTRODUCTION

Business Intelligence (BI) was introduced by the Gartner group in the middle of 1990s. It refers to the knowledge possessed about the competitors, customers, business partners and functioning of the internal things which helps in taking better decisions. It integrates/ combines databases, architectures, tools, analytic tools, methodologies and applications. Generally, it conveys different meaning to different people. Thus, it is known as a context-free expression. The main objective or goal of the Business Intelligence (BI) is to allow interactive access to the data in order to perform the manipulation of data. Besides this, it also focuses on providing suitable analysis to the business analysts and managers.

The architecture of Business Intelligence (BI) system consists of four major components. They are Data Warehouse, Business Analytics, Business Performance Management(BPM) and User Interface.

1.1 DEFINITION, HISTORY AND EVOLUTION

Q1. What is Business Intelligence? Discuss in detail about its history and evolution.

Answer :   *Model Paper-I, Q6(a)*

Business Intelligence (BI)

Business Intelligence (BI) refers to the Knowledge one possess about the competitors, customers, business patterns and functioning of the internal things Which helps them in taking better decisions.

Business Intelligence integrates databases, architectures, tools, analytic tools, methodologies and applications. Generally, it convey different meaning to different people. Thus, it is known as a context-free expression. The main objective or goal of the Business Intelligence (BI) allows interactive access to the data inorder to perform the manipulation of data. Besides this, it also focuses on suitable analysis to the business analysts and managers. Decision makers will take more knowledgeable and valuable decisions by analyzing the historical data, current data, performances and situations. The business intelligence (BI) process is dependent of various stages including the transformation of data into information, information to decisions and decisions to actions.
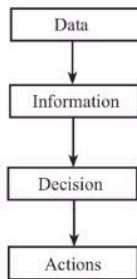
```
        ┌─────────────┐
        │    Data     │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │ Information │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │  Decision   │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │   Actions   │
        └─────────────┘
```

Figure: Transformations in business Intelligence process

History of BI

Business Intelligence (BI) was introduced by the Gartner group in the middle of 1990s. Even though the term and concept is old, it has its own importance in the Management Information System (MIS) reporting system of 1970s. Initially, the reporting systems were static two dimensional reporting and does not possess any analytical capabilities. Then, during 1980's, the concept of executive information system (EIS) is introduced that enhances the computerized support to the high-level executives and managers. This support include capabilities like dynamic and multi-dimensional reporting, trend analysis, forecasting and prediction, status access, drill drown to details and critical source Factors (CSF). Hence, set of several products are released with the provided capabilities. Therefore, the present capabilities and some additional capabilities of original EIS are transformed to BI. Later, during the period 2005, BI adds the additional capabilities of powerful analytic and artificial intelligence. Hence, the BI possess all the information that is required by the executives and managers.

Evolution of BI

The various techniques and tools in the BI-based enterprise information system that explains the evolution of BI, are shown in the below figure,
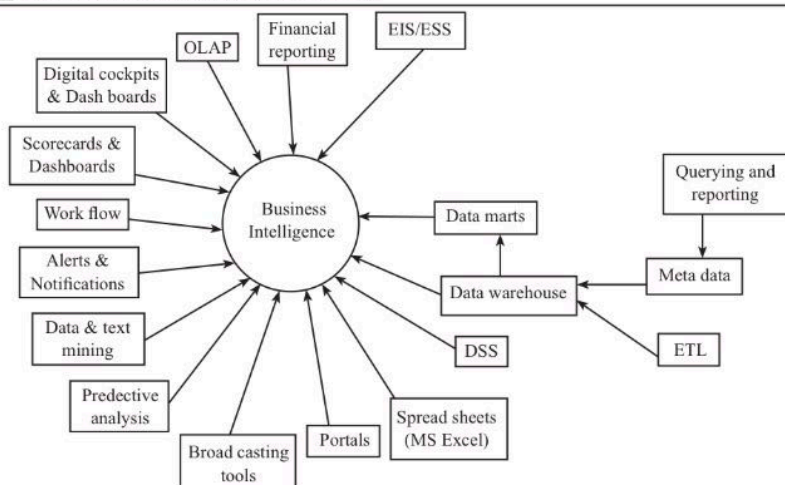
SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD. ───────────

Figure: Evolution of Business Intelligence (BI)

1.2 STYLES OF BUSINESS INTELLIGENCE, BENEFITS OF BUSINESS INTELLIGENCE

Q2. List the styles of business intelligence and also discuss its benefits.

Answer :   *Model Paper-II, Q6(a)*

Styles of Business Intelligence

The common styles of business intelligence are categorized into five wherein each style type is provided with certain specific tools. They are,

1. Enterprise reporting
2. Cube analysis/slice-and-dice analysis
3. Statistics and data mining
4. Report delivery and alerting
5. Ad hoc queries.

Benefits of Business Intelligence

Some of the benefits of business intelligence are as follows,

1. It is used in various applications such as general reporting, sales and market analysis, financial consolidation, planning and forecasting, budgeting, statutory reporting and profitability analysis.
2. It provides efficient and exact information that includes the real-time view of the corporate performance and the elements required for all types of decisions, strategic planning as well as survival.
3. It provides an efficient decision-making process so as to improve various business operations.
4. It provides an efficient, faster, more informed decision making (i.e. competitive imperative).
5. It provides an increasing win rate to the organizations.
6. It provides manager's with right information at the right time and in right place.
7. It is used in many typical analytic applications such as customer segmentation, customer profitability, customer attrition, fraud detection, channel optimization and so on.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

Q3. Discuss in detail about Automated Decision systems (ADS).

Answer : *Model Paper-III, Q6(b)*

Automated Decision System (ADS) is a new rule-based system that is introduced to support decision-making process. so, it is also called as Decision Automated System (DAS). It provides solution to various problems such as one functional area to a certain repetitive managerial problem, industry problem (i.e. pricing of the product and services).

The following figure describes the framework of automated decision-making system.
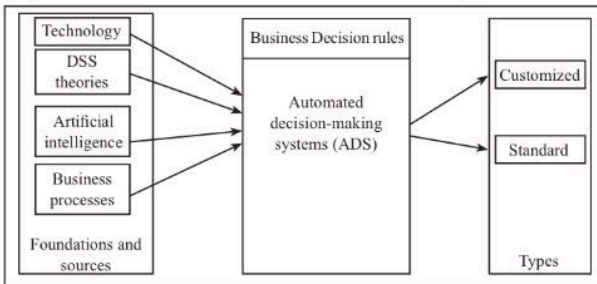

Figure: Automated Decision-making Framework

Initially, the automated decision systems are used in airline industries. In airlines industries, ADS are used to price the tickets dynamically depending up on the actual demand. They provides rule-based solutions rather than the various management science approaches that provides model based solutions to generic structured problems. Hence, they are known as revenue management or revenue optimization systems.

Some of the examples of business rules included in this rule-based systems are as follows,

(i) It provides a discount of 'p' to the non-business passengers only when 80% of seats in the flight from source (i.e. sydney) to destination (i.e., Colambia) are sold out before 4 days to departure.

(ii) A user can be provided with $20,000 credit line, when the user has own house and earns more than $2,00,000 per year.

(iii) If the cost of a product is more than $2,000 then the company or organization owns that product only once in a year, and the purchasing agent does not requires any specific approval.

The above mentioned examples of business rules that are derived from the real-time experiences [or through the datamining] can be integrated with mathematical model and generates the solution to the problems, or provided to human for finalizing the decisions. ADS also tries to automate highly repetitive decisions depending on the business rules. This type of systems are efficient and suitable for frontline systems, who can able to view the information of a customer and provides quick decisions in a particular periodic of time. Event-driven alert is an example of ADS.

1.3 REAL-TIME BUSINESS INTELLIGENCE

Q4. Discuss in detail about real-time business intelligence.

Answer : *Model Paper-III, Q6(a)*

Business users (Application managers, role managers and risk managers etc.) are progressively demanding access to the real-time data (or unstructured data) that includes the contents of data warehouse. For instance, the buses in foreign countries like Texas, Houston are designed with several facilities like data gathering devices, traffic controllers that are capable of accessing and changing the traffic signal/lights interval. Thus, making the system more efficient and reliable. An organization is said to be successful (or survive) only when it deals with the real-time data updates. For instance, the cities nearer to the sea and with low elevation distance make use of the real-time data collection and analysis. By using this equipment the officials will get to know the present situations of the cities, like the working of the flood-controlling pumps which takes an active participation in saving the people when the floods occurs. Interestingly, it is also used in several parts of the world like Netherlands and New Orleans.

Earlier, BI tools and dataware housing mainly focuses on assisting managers in preparing tactical and strategic decisions with the help of the historical data. In the year 2003, a shift has started using real-time data warehousing for making operational decisions. Some of the IT vendors like Microsoft, Oracle, IBM and other are preferring to actively use these technologies. Hewlett-packard has turned towards an adaptive enterprise strategy to deliver the on-demand computing.

Today, there is a faster development in the business intelligence (BI) software which generates the real-time data for the real-time business analysis and decision making. This growth includes acquiring the right information to operational and tactical personnel who are responsible for managing the short-term features to run an organization. Thus, they can use new business analysis and up-to-the-minute results for taking decisions.

Now a days, the demand of the current data is more. So, most of the IT manager have queries like how the real-time business analysis can be conducted, etc. Many of the real-time projects are in development and deployment stages only and the demand towards the real-time application will continues to increase. For instance, the growth of business rules management and ADS, enforces to implement most of the automated business processes in data ware house environment. The reply of a real-time system is mandatory when the processes need immediate data updates then it required to answer the analytical queries. Thus, the response time for data mining, OLAP and queries is approximately zero.

Real-time business analysis can be achieved by more frequently updating the data warehouses. In the year 2003, the daily basis updation of data warehouses became famous and the time interval is continues to decrease. Moreover, business analysis applications are being deployed apart from real-time queries. For instance, such application can quickly recognize, customer buying patterns on the basis of store displays, provides an instantaneous modifications for replacement or display itself. Some of the other applications are fraud detection, revenue management, call-centre support transportation and other financial transactions.

Real-time requirement includes various queries that updates the data. Hence, these requirements can also change the design view of a database, OLAP, data warehouses and several data mining tools. Alternatively , substantial business value demonstration can be adopted for a business process. Some of the examples of web-based and real-time business intelligence softwares are as follows,

(i)   Live Business (Data mirror)

(ii)   Live sheet for Excel (KnowNow)

(iii) NetIQ corporation

(iv)   Supply chain analytics and BI series 8(cognos).

(v)   Web intelligence (business objects)

(vi)   DB2 intelligent minor scoring (IMS; IBM)

(vii) Databeacon (Databeacon Inc.)

(viii) SonicMQ (Sonic Software)

(ix)   Supply chain Intelligence Suite (SAS)

(x)   Power Center standard and Advanced versions (Informatica).

A major issue in real-time system is that the data in the data warehouse is not continuously updated. This leads to the occurrence of problems while dealing with the real-time data and in its report generation. The real-time data is required while constructing the ADS systems.

1.4 BUSINESS INTELLIGENCE VALUE CHAIN

Q5. Write about business intelligence value chain.

Answer :   *Model Paper-I, Q6(b)*

The Business intelligence value chain mainly focuses on the business goals of an organisation. A value chain system comprises of the following,

(i)   Data mining

(ii)   Data visualization

(iii) Real-time dataware housing

(iv) Automatic learning and refinement

(v)   Automated anomaly and exception detection.

(vi) Geographic information systems

(vii) Seamless flow-through workflow

(viii) Proactive alerting with automatic recipient determination.

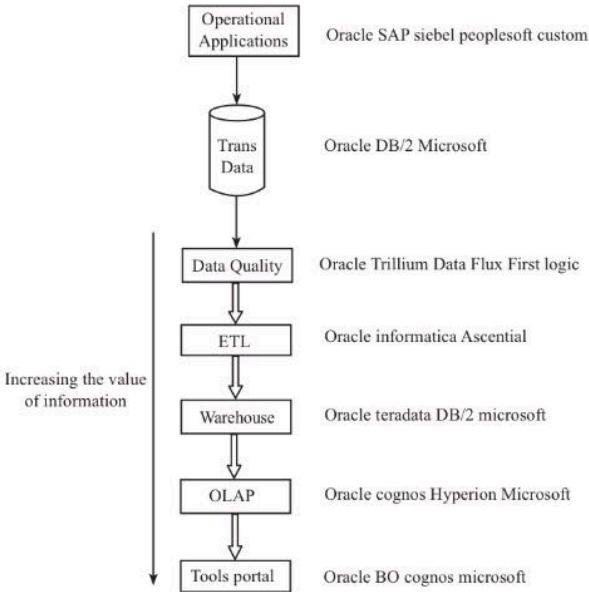The following figure describes the business intelligence value chain.

Operational Applications	Oracle SAP siebel peoplesoft custom
Trans Data	Oracle DB/2 Microsoft
Data Quality	Oracle Trillium Data Flux First logic
ETL	Oracle informatica Ascential
Warehouse	Oracle teradata DB/2 microsoft
OLAP	Oracle cognos Hyperion Microsoft
Tools portal	Oracle BO cognos microsoft

Increasing the value of information

Figure: Business Intelligence Value Chain

The data source can be added with a new value in every step of the business intelligence value chain. The process involved in the business intelligence value chain is,

Data resources → Information → knowledge worker → Business intelligence → Business strategies → Business goals

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

1. Data resource helps in generating the information using the information using the information engineering process.

2. Information helps in generating the knowledge worker in a knowledge environment.

3. Knowledge worker helps in generating business intelligence in the intelligent learning enterprise.

4. Business intelligence helps in generating the business strategies.

5. Business strategies helps in the generation of the goals of an enterprise or organization. Which is the last step in the business intelligence value chain.

1.5 ARCHITECTURE OF BUSINESS INTELLIGENCE

Q6. Explain the architecture of business intelligence.

Answer :   *Model Paper-II, Q6(b)*

The architecture of Business Intelligence (BI) system consists of four major components. They are,

1. Data Warehouse

2. Business Analytics

3. Business Performance Management (BPM)

4. User Interface.

The following figure describes the high-level architecture of business intelligence (BI) and also the relationship among the four major parts of business intelligence.
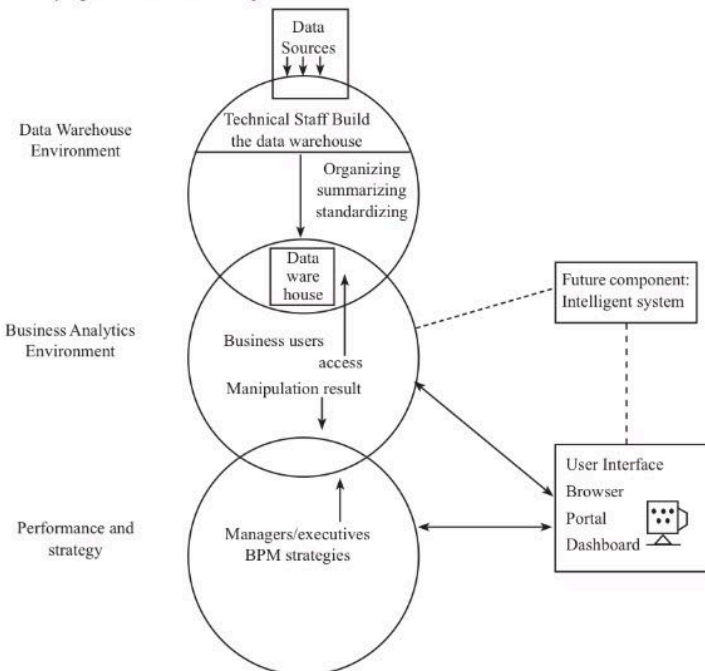

Figure: High-level Architecture of Business Intelligence (BI)

1. **Data Warehouse**

The data warehousing is mainly deals with the technical staff. Data warehouse and its elements are known to be as the foundation of medium-to-large business intelligence system. Generally, it consists of organized & outlined historical data that helps the end users to access the data and information easily. Now a days, some of the data warehouses include historical data as well as the current data, there by offering the real-time decision support.

2. **Business Analytics**

The data or information in a data warehouse can be accessed by the end users through various tools and techniques. These tools and techniques are divided in to two categories. They are,

(i) **Data, Text, Web Mining and other Sophisticated Mathematical and Statistical Tools**

The main purpose of data mining is to discover the relation that connects different database entities. The process of data mining can be done on textual data and also on web data. Data mining can be done by using various tools and techniques such as predictive analysis-techniques, neural computing, artificial intelligence techniques, statistical and mathematical techniques.

(ii) **Reports and Queries**

Business analytics consists of various reports and queries such as, static and dynamic reporting, multi-dimensional view, drill down to details, discovery of information and so on.

3. **Business Performance Management (BPM)**

Business performance management is also called as corporate performance management. BPM includes the evolution and architecture of Business Intelligence (BI).

BPM introduced a new concept called as management and feedback that extends the measuring, monitoring and comparison of sales, cost, profit and profitability. It involves in various processes such as planning, forecasting and budgeting. The traditional Decision Support Systems (DSS), Executive Information System (EIS), Business Intelligence (BI) helps in the bottom-up extraction of information from data where as BPM provides/offers a top-down application of corporate wide strategy. Moreover, it is basically combines with the balanced scorecard methodology as well as dashboards.

4. **User Interface**

User interface includes dashboards and other information broadcasting tools such as dash boards, portal and browser. Using these tools, a user can connect to user interface. Dashboards are meaningful groups of corporate/marketing performance measures(key performance indicators), exceptions and trends. They are used to combine the information collected from various business areas and also provides a graph that shows the comparison between the actual performance and the required metrics. Hence, it shows view of the organization's health. In addition to this, user interface includes some other information broadcasting tools such as digital cockpits, corporate portals and visualization tools that ranges from multidimensional cube presentation to virtual reality are known to be the integral parts of business intelligence (BI) systems. Since, BI is derived from EIS, BI acquires many visual aids for executives are converted into BI software. However, Geographical Information system (GIS) also plays an incremental role in DSS.

SHORT QUESTIONS AND ANSWERS

Q1. Write a short note on business intelligence.

Answer :   *Model Paper-I, Q1*

Business Intelligence refers (1) to the Knowledge one possess about the competitors, customers, business patterns and functioning of the internal things. Which helps them in taking better decisions.

Business Intelligence integrates databases, architectures, tools, analytic tools, methodologies and applications. Generally, it convey different meaning to different people. Thus, it is known as a context-free expression. The main objective or goal of the Business Intelligence (BI) allows interactive access to the data inorder to perform the manipulation of data. Besides this, it also focuses on suitable analysis to the business analysts and managers. Decision makers will take more knowledgeable and valuable decisions by analyzing the historical data, current data, performances and situations. The business intelligence (BI) process is dependent of various stages including the transformation of data into information, information to decisions and decisions to actions.

Q2. Describe the history of BI.

Answer :

Business Intelligence (BI) was introduced by the Gartner group in the middle of 1990s. Even though the term and concept is old, it has its own importance in the Management Information System (MIS) reporting system of 1970s. Initially, the reporting systems were static two dimensional reporting and does not possess any analytical capabilities. Then, during 1980's, the concept of executive information system (EIS) is introduced that enhances the computerized support to the high-level executives and managers. This support include capabilities like dynamic and multi-dimensional reporting, trend analysis, forecasting and prediction, status access, drill drown to details and critical source Factors (CSF). Hence, set of several products are released with the provided capabilities. Therefore, the present capabilities and some additional capabilities of original EIS are transformed to BI. Later, during the period 2005, BI adds the additional capabilities of powerful analytic and artificial intelligence.

Q3. List the benefits of BI.

Answer :   *Model Paper-I, Q1*

Some of the benefits of business intelligence are as follows,

1. It is used in various applications such as general reporting, sales and market analysis, financial consolidation, planning and forecasting, budgeting, statutory reporting and profitability analysis.

2. It provides efficient and exact information that includes the real-time view of the corporate performance and the elements required for all types of decisions, strategic planning as well as survival.

3. It provides an efficient decision-making process so as to improve various business operations.

4. It provides an efficient, faster, more informed decision making (i.e. competitive imperative).

5. It provides an increasing win rate to the organizations.

6. It provides manager's with right information at the right time and in right place.

7. It is used in many typical analytic applications such as customer segmentation, customer profitability, customer attrition, fraud detection, channel optimization and so on.

Q4. Discuss some of the examples of business rules included in rule-based systems.

Answer :

Some of the examples of business rules included in this rule-based systems are as follows,

(i)   It provides a discount of 'p' to the non-business passengers only when 80% of seats in the flight from source (i.e. sydney) to destination (i.e., Colambia) are sold out before 4 days to departure.

(ii) A user can be provided with $20,000 credit line, when the user has own house and earns more than $2,00,000 per year.

(iii) If the cost of a product is more than $2,000 then the company or organization owns that product only once in a year, and the purchasing agent does not requires any specific approval.

The above mentioned examples of business rules that are derived from the real-time experiences [or through the datamining] can be integrated with mathematical model and generates the solution to the problems, or provided to human for finalizing the decisions.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

Q5.   List the components of business value chain.

Answer :

A value chain system comprises of the following components,

(i)   Data mining

(ii)   Data visualization

(iii) Real-time dataware housing

(iv)   Automatic learning and refinement

(v)   Automated anomaly and exception detection.

(vi)   Geographic information systems

(vii) Seamless flow-through workflow

(viii) Proactive alerting with automatic recipient determination.

Q6.   Discuss the process of BI value chain.

Answer :

The process involved in the business intelligence value chain is,

Data resources $\rightarrow$ Information $\rightarrow$ knowledge worker $\rightarrow$ Business intelligence $\rightarrow$ Business strategies $\rightarrow$ Business goals

1.   Data resource helps in generating the information using the information using the information engineering process.

2.   Information helps in generating the knowledge worker in a knowledge environment.

3.   Knowledge worker helps in generating business intelligence in the intelligent learning enterprise.

4.   Business intelligence helps in generating the business strategies.

5.   Business strategies helps in the generation of the goals of an enterprise or organization. Which is the last step in the business intelligence value chain.

Q7.   What are the tools and techniques involved in business analytics?

Answer :   *Model Paper-III, Q1*

The data or information in a data warehouse can be accessed by the end users through various tools and techniques. These tools and techniques are divided in to two categories. They are,

(i)   Data, Text, Web Mining and other Sophisticated Mathematical and Statistical Tools

The main purpose of data mining is to discover the relation that connects different database entities. The process of data mining can be done on textual data and also on web data. Data mining can be done by using various tools and techniques such as predictive analysis-techniques, neural computing, artificial intelligence techniques, statistical and mathematical techniques.

(ii)   Reports and Queries

Business analytics consists of various reports and queries such as, static and dynamic reporting, multi-dimensional view, drill down to details, discovery of information and so on.

Q8.   Write a brief note on user interface component of BI.

Answer :

User interface includes dashboards and other information broadcasting tools such as dash boards, portal and browser. Using these tools, a user can connect to user interface. Dashboards are meaningful groups of corporate/marketing performance measures(key performance indicators), exceptions and trends. They are used to combine the information collected from various business areas and also provides a graph that shows the comparison between the actual performance and the required metrics. Hence, it shows view of the organization's health. In addition to this, user interface includes some other information broadcasting tools such as digital cockpits, corporate portals and visualization tools that ranges from multidimensional cube presentation to virtual reality are known to be the integral parts of business intelligence (BI) systems. Since, BI is derived from EIS, BI acquires many visual aids for executives are converted into BI software. However, Geographical Information system (GIS) also plays an incremental role in DSS.

INTERNAL ASSESSMENT

I. **Multiple Choice**

1. _____ is the type of styles of BI.   [   ]

    (a) Enterprise reporting

    (b) Statistics and data mining

    (c) Ad hoc queries

    (d) All the above

2. In the first step of BI process, data is transformed into _____.   [   ]

    (a) Information

    (b) Decision

    (c) Action

    (d) None of the above

3. Introduction of _____ system enhanced the computerized support to high level executives and
    managers.   [   ]

    (a) MIS

    (b) EIS

    (c) Multi-dimensional reporting

    (d) CSF

4. _____ provides rule-based solution to generic structured problems.   [   ]

    (a) MIS

    (b) ADS

    (c) EIS

    (d) CSF

5. _____ is not a web-based and real-time business intelligence software.   [   ]

    (a) Web intelligence

    (b) SonicMQ

    (c) Enterprise reporting

    (d) Databeacon

6. _____ is a component of value chain system.   [   ]

    (a) Data mining

    (b) OLAP

    (c) Business analytics

    (d) All the above

7. _____ helps in generating information using information engineering in BI value chain.   [   ]

   (a)   Information

   (b)   Knowledge worker

   (c)   Data resource

   (d)   Business strategy

8. What is the last step of BI value chain?   [   ]

   (a)   Information

   (b)   Knowledge worker

   (c)   Data resource

   (d)   Business strategy

9. _____ deals with technical staff in BI architecture.   [   ]

   (a)   Data warehousing

   (b)   Business analytics

   (c)   Reports and queries

   (d)   User interface

10. The main purpose of _____ is to discover relation that connects different database entities.   [   ]

   (a)   Data warehousing

   (b)   Data mining

   (c)   BPM

   (d)   User interface

II.   Fill in the Blanks

1. Business intelligence is also known as _____.

2. The latest technology used in BI is _____.

3. The common styles of BI are categorized into _____ types.

4. ADS stands for _____.

5. Event-driven alert is an example of _____.

6. The role of business intelligence in BI value chain is to generate _____.

7. Hawlett-Packard(HP) turned towards an adaptive enterprise strategy to deliver _____.

8. BPM is also called as _____.

9. _____ are meaningful groups of corporate/marketing performance measures, exceptions and trends.

10. _____ includes dashboards are other information broadcasting tools.

KEY

I.   Multiple Choice

1. (d)

2. (a)

3. (b)

4. (b)

5. (c)

6. (a)

7. (c)

8. (d)

9. (a)

10. (b)

II.   Fill in the Blanks

1. Content-free expression

2. Artifical Intelligence

3. Five

4. Automated Decision System

5. ADS

6. Business strategies

7. On-demand computing

8. Corporate performance management

9. Dashboards

10. User interface.

III. Very Short Question and Answers

Q1. Define Business Intelligence (BI).

Answer :

Business Intelligence (BI) refers to the Knowledge one possess about the competitors, customers, business patterns and functioning of the internal things which helps them in taking better decisions.

Q2. What is ADS?

Answer :

Automated Decision System (ADS) is a new rule-based system that is introduced to support decision-making process. so, it is also called as Decision Automated System (DAS). It provides solution to various problems such as one functional area to a certain repetitive managerial problem, industry problem (i.e. pricing of the product and services).

Q3. What is the major issue of real-time systems?

Answer :

A major issue in real-time system is that the data in the data warehouse is not continuously updated. This leads to the occurrence of problems while dealing with the real-time data and in its report generation.

Q4. What is the role of knowledge worker in BI value chain?

Answer :

The role of knowledge worker is to help in generating business intelligence the intelligent learning enterprise.

Q5. What is included in the queries and reports of business analytics?

Answer :

Business analytics consists of various reports and queries including static and dynamic reporting, multidimenstional view, drill down to details, discovery of information and so on.

UNIT 2

Data Warehousing and Data Mining

INTRODUCTION

Data warehouse is a form of storage system (database) where large volume of data is stored in such a way that retrieving desirable information from the system becomes easy and reliable. W.H. Inmon defined data warehouse as a subject (but not application) oriented, consolidated, time-dependent and non-erasable collection of data which is mainly supported by decision management. There are three main types of data warehouses. They are data marts, operational data stores (ODS) and enterprise data warehouses (EDW). It can be visualized as 3-tier architecture which is a preferable for fulfilling the large scale and performance specific data requirements.

Real-time data warehousing is a type of data warehousing in which the data is processed and loaded instantly when it becomes available. It is also called as active data warehousing. It transforms the functions of data warehouse into tactical decision making functions. Using this type of data warehousing, all the individuals of the organization who interact with the customers are provided with information based decision making.

Data mining is a process of extracting knowledge from massive volumes of data. It refers to a way of finding significant and useful information from an organization's database. The knowledge which is extracted can include pattern types, association rules and different trends. Data mining is not confined to a particular organization, instead, it has techniques to explore the knowledge hidden in any data. The different techniques used for digging out data are artificial intelligence, statistical and mathematical techniques and pattern recognition techniques. The tasks or functions of data mining are classified into prediction, association and clustering. Some of the software tools of data mining are IBM intelligent miner, SPSS Clementine, Oracle data mining, classification and regression tree and weka.

2.1 DATA WAREHOUSING (DWH)

2.1.1 Definition, Characteristics, Types

Q1.   What is a data warehouse? Discuss its characteristics.

Answer :   *Model Paper-I, Q7(a)*

Data Warehouse

Data warehouse is a form of storage system (database) where large volume of data is stored in such a way that retrieving desirable information from the system is very easy and reliable. Data warehouse is stored in different location so that it doesn't collide with transactional database system which stores day-to-day information and answers the queries that are prerecorded in the database. Data warehouse system on the other hand provides solutions to sophisticated queries, which involves many computations to be performed at finer-level of granularity.

W.H. Inmon defined data warehouse as a subject (but not application) oriented, consolidated, time-dependent and non-erasable collection of data which is mainly supported by decision management. From this definition, the following characteristic features of data warehouse can be defined as,

    1.   Subject-oriented (Not application-oriented)

    2.   Consolidated data

    3.   Time-dependent data

    4.   Non-erasable data.

1.   Subject-oriented

Data warehouse focuses mostly on important subjects as operational databases like producer, consumer, manufacturer i.e., data warehouse in not application-oriented. The major role of data warehouse is to support decision makers in making strategic decisions. It does this by performing data analysis and by applying data modelling tools. Data warehouse abstract unnecessary data which is not required in decision making and provide a precise view of data organized around a specific subject. Data warehouse doesn't maintain information about day-to-day transactions of organization but concentrate on the subjects which are critical to the organization.

2.   Consolidated Data

Data warehouse is capable of retrieving appropriate data from heterogeneous databases (like relational databases, flat files) in order to make efficient strategic decisions. Because of the heterogeneity, data is stored in inconsistent manner. To confirm consistency and reliability in naming rules, encoding methods, different techniques such as data cleaning and data integration needs to be applied. Therefore, it is necessary to perform data transformation, data consolidation before transferring data from operational system into data warehouse.

3.   Time-dependent Data

Data warehouse database not only stores current information but also stores historic information about a particular transaction. In operational database, though historic information is stored, it generates only present information as these databases are capable of supporting only day-to-day transactions.

Data in data warehouse are archieved as snapshots over historic and present time periods. In these databases, every data structures are time dependent i.e., they are directly or indirectly elements of time. This data warehouse approach is non-trivial for design as well as for implementation stages of data warehouse.

The advantages due to time-dependent feature in data warehouse are,

   ❖   It enables analyzing of historic data.

   ❖   It associates information of past data with present data.

   ❖   It provides better future prediction.

4.   Non-erasable Data

This feature confirms that once data enters the data warehouse it remains static until particular event is triggered. Data warehouse contains data which is extracted transformed, integrated from operational database. The data is transferred from operational system to data warehouse at regular intervals of time depending on the specification of the business. There are only two operations executed by data warehouse for accessing data.

   (a)   Data loading method

   (b)   Data access method.

Because of the above features, data warehouse can be considered as a consistent storage area that provides support for decision making and for analytical reporting.

Apart from these characteristics, data warehouse provides certain additional characteristics which are as follows,

❖   It can be used in web based applications to facilitate efficient computing platform.

❖   It can employ relational or multidimensional structure.

❖   It is designed to adopt client/server architecture which simplifies data access.

❖   Modern data warehouses are capable of providing real time access to the data.

❖   It provides metadata which describes the way in which the data is stored in the warehouse. This data can be used to quickly and efficiently access the stored data.

Q2.   Describe various types of data warehouses.

Answer :   *Model Paper-II, Q7(a)*

There are three main types of data warehouses. They are,

   1.   Data marts

   2.   Operational Data Stores (ODS)

   3.   Enterprise Data Warehouses (EDWs).

1.   Data Marts

Data marts contain versatile set of data which is selected from enterprise wide data. The data is selected based on the requirement of particular user group. It represents data from a single process of business. The difference between enterprise warehouse and data mart is that, the former is cross-functional in scope, but the latter is limited to particular selected subjects. The data in data marts are not in detailed format but in summarized and simplified form.

Data marts are divided into two-types. They are,

   (a)   Independent data mart

   (b)   Dependent data mart.

(a)   Independent Data Mart

These marts are small and low cost version of data warehouse. It usually contains data concerning to a particular department of the company. The data within this data mart is implemented with no effect or minimum effect on organizational transaction database.

(b)   Dependent Data Mart

These marts contain information taken explicitly from enterprise data warehouses. These marts focus on consistency and quality of data. They incur higher cost than independent databases.

2.   Operational Data Stores (ODS)

The operational data stores carry the most recently updated information regarding the customers. This information is updated throughout the life cycle of a business operation. Typically, the operations to which this information belongs are short term involving decisions regarding applications designed for a specific purpose. When an update is found, the old file is replaced with new one in ODS. The information in ODS can be related to multiple sources which can provide near real-time data. The data of ODS acts as input to oper mart which is used to analyze the data multidimensionally.

3.   Enterprise Data Warehouses (EDWs)

All the information associated with the important subjects that is dispersed across the whole enterprise is present in the enterprise data warehouse. The storage space of data can vary from few giga bytes to hundreds of tera bytes. Because of huge space, it includes descriptive as well as simplified data. These warehouses are realized on conventional large speed computers, super servers etc. The advantage of this warehouse is that, it allows enterprise-wide integration of data typically from at least one transaction system or from exterior information providers. The drawback is that a wide spread business modeling is needed. A huge span of time (in years) is consumed for designing and building these warehouses.

Q3.   Discuss in brief about metadata.

Answer :

Metadata can be defined as the data about data. It describes the data warehouse architecture by providing different views, dimensions, definitions, data mart locations, warehouse schema etc. In other words, metadata describes how the data is structured or stored in a warehouse. The definition of metadata is based on technical or business metadata usage. It can also be represented in the form of pattern view with which different types of data like syntactic data, structural data and semantic data can be separated.

Metadata patterns can be implemented using a strategy that follows a holistic approach that provides integration of enterprise metadata.

This strategy involves the following,

1.   Data extraction

2.   Data transformation

3.   Enterprise Information Integration (EII)

4.   Registries of metadata

5.   Ontology

6.   Service Oriented Architecture (SOA).

These concepts are involved to fulfill the following requirements,

1.   Improving effectiveness

2.   Increasing reusability

3.   Improving efficiency

4.   Achieving interoperability

5.   Providing flexibility

6.   Decreasing maintenance cost

7.   Providing user interface

8.   Improving overall performance.

Business metadata include information that is capable of generating meaningful information that helps in gaining the knowledge about the structure of data in the warehouse. Therefore, it can be said that business metadata converts the data/ information present in the warehouse to knowledge. According to zhao, metadata management maturity can be describe in five different levels. They are,

1.   Ad hoc

2.   Discovered

3.   Managed

4.   Optimized

5.   Automated.

2.1.2 Data Warehousing Framework

Q4. Give an overview of data warehousing process framework.

Answer :

Data warehousing is important for most or all the organizations to manage large amount of data for assisting decision support system. Typical framework of data warehousing process is shown in the below figure.
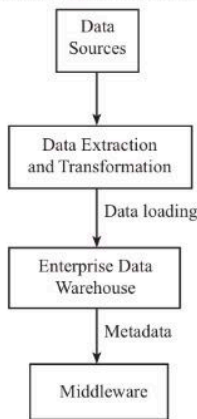

Figure: Data Warehousing Process Framework

The following are the components involved in data warehousing process framework.

1.   Data Sources

The data to be stored in a data warehouse can be extracted from many data sources such as ERP system, OLTP, web/internet, legacy system or any other source.

2.   Data Extraction and Transformation

The data is extracted from the data sources and it is converted into appropriate format using ETL (Extraction, Transformation and Loading) software.

3.   Data Loading and Enterprise Data Warehouse

The converted data is loaded into a comprehensive database called Enterprise Data Warehouse (EDW) which provides detailed information of data to support decision making.

4.   Metadata

Metadata is also included for users and data administrators to efficiently access the data stored in the warehouse. This metadata provides rules followed in arranging and storing the database so that the required data can be easily retrieved even with web tools.

5.   Middleware

Middleware offers tools which can be used for accessing the data. These tools can vary for different types of users. Some of the tools are data/text mining, OLAP, visualization tools. System analysts can access the data using Structured Query Language (SQL).

This process is followed in order to manage the complexities involved in maintaining and using massive amount of data present in a warehouse. It continuously provides reliability and availability even if the number of users increases anonymously.

2.1.3 DWH 3Tier Architecture

Q5.   **Explain in detail about 3-tier architecture of data warehousing.**

Answer :

Data warehouse can be visualized as 3-tier architecture which is a preferable architecture for fulfilling the large scale and performance specific data requirements.

The three tiers in this architecture are as follows,

Tier-1

This tier represents the front-end client layers. It carries a decision support system or business intelligence system along with the client who uses these systems to access the data in the warehouse.

Tier-2

This tier represents the application server in which data acquisition software resides. This software is responsible for extacting the data from various sources and storing it into the warehouse after transformation.

Tier-3

This tier represents warehouse server where all the data is stored along with its respective software. It is usually a relational database system where data is stored in the form of tables.

The typical architecture of a 3-tier data warehouse is shown in the figure below,
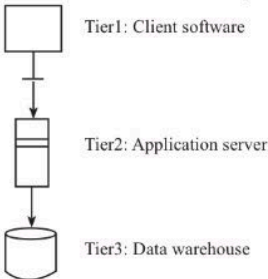
Tier1: Client software

Tier2: Application server

Tier3: Data warehouse

Figure: 3-tier Architecture of Data Warehouse

The data present in the warehouse can be further processed and stored in data marts or multidimensional database to simplify the retrieval or analysis. The above architecture helps in creating data marts easily and removing restrictions related to resources. This is because it separates the various functionalities of data warehouses. It also provides improved performance when compared with two tier architecture.

In web-based data warehousing, the two technologies i.e., internet and data warehousing are integrated. This type of data warehousing can also be visualized as a 3-tier architecture as shown in the figure below.
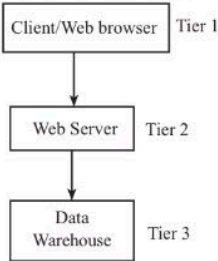
Client/Web browser   Tier 1

Web Server   Tier 2

Data Warehouse   Tier 3

Figure: 3-tier Web-based Data Warehousing Architecture

The three tiers in this architecture are as follows,

Tier-1

This tier represents the front-end client layer. The client connects with the webserver through a web browser connected over internet, intranet or extranet. Web browsers are enabled with Graphical User Interface(GUI).

Tier-2

This tier represents web server which is responsible for handling the requests generated by the clients to access data from the warehouse.

Tier-3

This tier represents data warehouse which enables the web server to access data from it.

It is necessary to consider the speed of loading web pages on the web browser which can be managed by planning the server capacity.

Use of web based 3-tier architecture offers the following advantages,

❖   It minimizes the overall cost

❖   It makes the data access platform independent

❖   It simplifies the data access.

Q6.   List the issues involved in selecting among DWH architectures.

Answer :

Following are the issues involved in selecting among various DWH architectures.

1.   Choice of Selecting DBMS

Data warehouses usually employ RDBMS (Relational Database Management Systems) as database management system. However, some warehouses also employ SQL, Oracle and DB2. All these systems are capable of working with client/server and web based architectures.

2.   Choice of Selecting Among Parallel Processing and Partitioning

Parallel processing approach can be used to allow more than one computer to parallely process the queries. Use of such an approach makes the system scalable. In order to provide efficiency, designers must decide whether to partition or split the database or not. This is a major issue when massive amount of data is stored in a warehouse.

3.   Choice of Whether to Use Data Migration Tools or Not

Data migration refers to the moving of data from one system to another system or a warehouse. It depends on different factors such as location of data assets and diversity. Depending on these factors, designers need to decide whether to employ various migration tools or not.

4.   Choice of Selecting Data Retrieval and Analysis Tools

Data retrieval and analysis require specific tools for data to be extracted, processed, transformed and loaded in the data warehouse. These tools can be created by designers, purchased from a provider or an existing tool can be used.

2.1.4 Alternative Architectures

Q7.   Discuss various alternative architectures of data warehousing.

Answer :

The alternative architectures of data warehousing are as follows,

1.   Independent Data Marts Architecture

In this architecture individual data marts are created which works without depending on each other. These marts focus on individual goals instead of fulfilling the overall organizational goals. It is the simplest and cost effective architecture among all the other architectures. The drawback of this architecture is that there exist data inconsistencies leading to complexities in analyzing data present in different data marts.
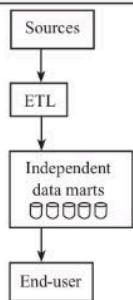
Figure: Architecture of Independent Data Marts

2.   Data Mart Bus Architecture

This architecture is an extension of independent data marts architecture. In this architecture, the data marts are connected together through a middleware. Using such an architecture data consistency can be achieved. It can process complex queries but does not fulfil the performance requirements.
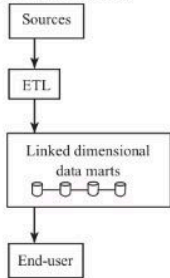

Figure: Architecture of Data Mart Bus

3.   Hub-and-Spoke Architecture

The main objective of this architecture is to achieve scalability and maintainability within the infrastructure. To achieve this, different subject areas are considered one after the other in an iterative way. It carries a centralized warehouse and multiple data marts which are interdependent. The advantage of this architecture is its simplicity and customization support for interfaces and reports. However, the drawbacks of this architecture includes redundancy and latency of data.
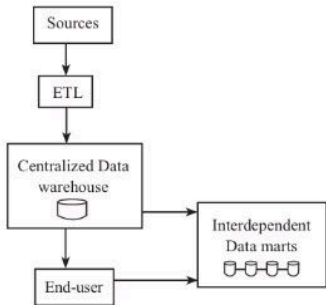

Figure: Architecture of Hub-and-spoke Data Warehousing

4.   Centralized Data Warehouse Architecture

In this architecture, a single data warehouse is used without any data marts. Such an architecture eliminates the restrictions on data which are imposed by data marts. In this way, the burden on technical team to classify or modify data according to different data marts gets eliminated.
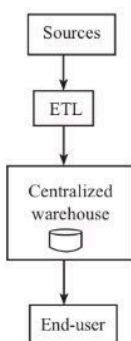

Figure: Architecture of Centralized Data Warehousing

5.   Federated Architecture

In this architecture, all the ways in which analytical resources can be integrated to fulfill the evolving business requirements are determined. This architecture integrates systems which are of different kinds and cannot be compared. It is typically employed by middleware tool vendors that support join and distributed queries. These tools are based on XML(Extensible Markup Language) which provides the overall view of distributed data sources. These architectures are used to avoid data warehouse replacement.
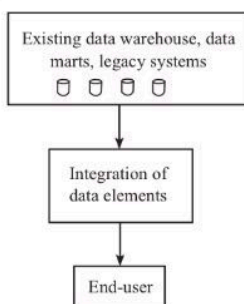

Figure: Federated Architecutre

2.1.5 Data Warehousing Integration

Q8.   Discuss in brief about data integration.

Answer :   *Model Paper-III, Q7(a)*

Data Integration

Data integration refers to the process of combining data from different data sources such as databases, flat files etc. It involves three processes which enable users to access data and also to array of ETL, analysis tools and data warehousing environment to access the data. These three processes are as follows,

1.   **Data Access**

This process allows data to be accessed and extracted from any type of sources.

2.   **Data Federation**

This process allows business views to be integrated with different data stores.

3.   **Change Capture**

This process captures the modifications done with respect to identification, capture and delivery of data sources.

Integration of data and metadata can be performed using the following technologies.

(i)   **Enterprise Application Integration (EAI)**

This technology helps in storing the data in the warehouse from various source systems. Integration is performed with respect to the functionality of application. Here, the functionality is distributed among multiple system to achieve reuse and flexibility. EAI can be used for different purposes such as forwarding decisions to OLTP, performing data acquisition on near-realtime data warehouse.

(ii)   **Service-Oriented Architecture (SOA)**

This technology helps in integrating business processes by dividing large application into smaller modules or services. It is used to maintain proper data flow in data warehouse which allows creation of a new function for each data element.

(iii)   **Enterprise Information Integration (EII)**

This technology uses a set of tools to ensure real-time data integration over a set of sources. Source can be a relational data base, multi dimensional database or web service. The tools offered by EII makes the data available to the users in the form of relational data by making use of views. The most critical component of this approach is XML, which provides various tags to generate knowledge at any stage.

4.   **Extraction, Transformation and Load (ETL)**

For answer refer Unit-II, Page No. 2.10, Q.No. 9.

Q9.   Describe the ETL process.

Answer :

ETL

ETL is a set of Extraction, Transformation and Loading processes which are performed in the data staging environment. It typically converts the irrelevant data extracted from several operational systems to the relevant/useful information. This information thus be stored into the data warehouse environment. The warehouse cannot be implemented efficiently if the data is not collected, cleansed and integrated in an appropriate way. Hence, the ETL process must be efficiently designed and implemented for optimized development of data warehouse.
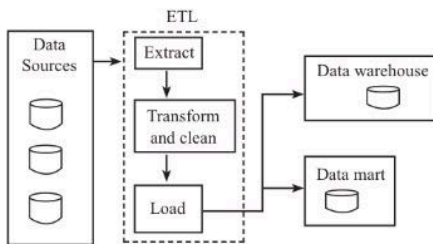

Figure: ETL Process

1. **Data Extraction**

The function of the data extraction is to find and extract the data from various relational and non-relational data sources for future use. Data extraction process includes,

(i)   Determining the operational systems from which the data can be collected.

(ii) Capture the appropriate data from data sources to make required data analysis.

(iii) Identify the extraction frequency of data i.e., daily, weekly or monthly to keep the data warehouse up to date.

2. **Data Transformation**

Data transformation is a process in which the data collected from many dissimilar source systems are transformed in accordance to a standardized format. The main purpose of performing a data transformation is to enhance the quality of the data before providing it as input in the data warehouse. Before transformation process, reliability, consistency and validity of the data are checked.

Data transformation is carried out by the following steps,

❖   Data elements should be identified that are to be moved to data warehouse.

❖   Calculated and derived data values should be identified.

❖   Next step is to perform data cleaning.

❖   Define the rules for transformation, data mapping criteria and metadata updation.

❖   In this step, data should be restructured.

❖   Data fields from various sources should be combined into one entity.

❖   For the captured data, data values, data types and field lengths should be standardized.

❖   Data integrity should be maintained within the data warehouse.

❖   Change the data according to the structure and format of data warehouse.

❖   Analysis of data warehouse can be made easy by simplifying and re-arranging the individual fields.

Some of the transformation processes are aggregation, normalization, union, join, enforcing keys etc.

3. **Data Loading**

Data loading is a process of loading the extracted and transformed data into the data warehouse repository.

The data loading usually, requires the data warehouse to be kept off-line such that no users can access the warehouse at the time of loading. Thus, a specified time must be assigned for loading the data. This can be done by fragmenting the data loading process into smaller number of portions such that, only few tables can be loaded concurrently.

Once the loading process has been completed, the loaded data need to be tested so as to ensure whether these loads have been correctly done in the data warehouse repository. In addition to this, a plan is designed to test the quality of the loaded records.

2.1.6 Data Warehousing Development Approaches

Q10. Discuss in brief about the development approaches of data warehouse. Compare them.

Answer :

Development Approaches of Data Warehouse

Data warehouses can be developed using either of the following two approaches.

1.   Inmon Model or EDW approach

2.   Kimball Model or Data Mart approach.

1. **Inmon Model or EDW Approach**

This development approach was suggested by Bill Inmon. Due to the efforts made by him in providing efficient data warehousing, he is called as the father of data warehousing. This approach follows a top-down method which utilizeds certain tools like entity-relationship diagrams which are based on relational database . It is also called as Enterprise Data Warehouse (EDW) approach because it focuses on the requirements of EDW. It does not include data marts by default EDW approach can be designed by the following method suggested by Murtaza.

2.   Kimball Model or Data Mart Approach

This development approach was suggested by kimball which works on the principle of "plan big, build small" statement. This method follows a bottom-up approach that involves data marts focussing on specific areas or departments of the organization. In this approach, the data marts are developed one after the other.

Comparison Between Inmon and Kimball's Approaches

Inmon's Approach		Kimball's Approach	
1.	It follows a top-down approach.	1.	It follows a bottom-up approach.
2.	Data warehouse architecture considers the overall organization.	2.	Data marts are used which consider various departments of the organization.
3.	It follows a complex method.	3.	It follows a simple method.
4.	The origin of this approach is spiral method.	4.	The origin of this approach is RDBMS.
5.	The physical design of this approach is thorough.	5.	The physical design of this approach is light.
6.	This approach include tools like ERD, DFD etc.	6.	This approach include dimensional modeling tools.
7.	It offers very low accessibility to the end-users.	7.	It offers higher accessibility to the end-users.
8.	The objective of this approach is to offer a technical solution which follows database methods.	8.	The objective of this approach is to simplify the data access so that end-users can directly access the database.

2.1.7 Real-time Data Warehousing

Q11. What is real-time data warehousing? Discuss its need.

Answer :

Real-time Data Warehousing

Real-time data warehousing is a type of data warehousing in which the data is processed and loaded instantly when it becomes available. For this reason, it is also called as active data warehousing. RDW/ADW transforms the functions of data warehouse into tactical decision making functions. Using this type of data warehousing, all the individuals of the organization who interact with the customers are provided with information based decision making. This feature improves most of the parts of organization such as logistics, supply chain management, customer service etc.

Need for Real-time Data Warehousing

Data warehouses are usually updated weekly. Due to this, the organizations cannot take decisions regarding the transactions or operations that require real-time or near real-time response. Moreover, inputs to the data warehouse have now became much faster and thus demanding the need for generating the decision quickly and some times instantly. To fulfill such requirements, real-time data warehousing is required. One of the major systems that require RDW is E-business/E-commerce.

Following are the problems that arise when RDW is not employed.

❖   Data in the warehouse is not always up to date that can lead to information mismatch.

❖   Organizations need to wait for some days or weeks to get most recent information.

❖   It becomes complex to store and manage metadata in sync when hub-and-spoke architecture is employed.

❖   When a single-centralized data warehouse is used, it increases the cost of maintaining, developing and securing systems.

❖   When updating is performed, batch load and processing power needs to be very high. In such cases, the processing becomes slow.

RDW provides a special repository in which integrated information is stored. This information helps strategic and decision support of the company. The data is updated as soon as the events occur from the OLTP systems.

The RDW can involve certain issues related to BI activities, data modeling, design of database, scalability and maintainability.

Q12. Compare traditional data warehousing and active data warehousing environments.

Answer :

Traditional Data Warehousing		Active Data Warehousing	
1.	Traditional warehousing can only support strategic decisions.	1.	Active warehousing can support both strategic and tactical decisions.
2.	Generation of results is difficult.	2.	Generation of results is done with operations.
3.	Data can be updated daily, weekly or monthly.	3.	Data needs to be updated instantly (within minutes).
4.	Few users can generate queries to the system concurrently.	4.	Thousands of users can generate queries to the system concurrently.
5.	It imposes several restrictions on reporting for verifying patterns and processes.	5.	It provides flexibility with ad hoc reporting and machine based modelling for identifying relationships.
6.	Users involved in this environment include knowledge workers, internal users and power users.	6.	Users involved in this environment include external users, operational staff and call centre.

2.2 DATA MINING

2.2.1 Definition, Characteristics, Benefits

Q13. Define data mining. List its characteristics and benefits.

Answer :

Data Mining

Data mining is a process of extracting knowledge from massive volumes of data. It refers to a way of finding significant and useful information from an organization's database. The knowledge which is extracted can include pattern types, association rules and different trends. Data mining is not confined to a particular organization, instead it has techniques to explore the knowledge hidden in any data. The different techniques used for digging out data are artificial intelligence, statistical and mathematical techniques and pattern recognition techniques.

Organizations that makes use of data mining techniques are benefited in their corresponding business area by identifying the significant trends and anomalies that were not possible to be detected by a human analyst. The association shows important knowledge about the database and entities present in the database. The purpose of data mining is to discover relation that connects different database entities.

Characteristics and Objectives of Data Mining

Following are the characteristics and objectives of data mining.

❖ Data mining cleanse the data to be stored in the warehouse.

❖ It works on a simple client/server or web-based information system architecture.

❖ It includes modern tools such as visualization tools which eliminate the unnecessary information included in the corporate files or records. To identify the existence of such information, synchronization and massaging is performed.

❖ Data mining user is usually the end user provided with certain additional (powerful) tools that help in retrieving the data quickly.

❖ Data from data mining tools can be easily integrated with spreadsheets and other softwares so that they can be accessed easily.

❖ It also offers parallel processing.

Q14. Give a simple taxonomy of data in data mining.

Answer :   *Model Paper-I, Q7(b)*

In a simple taxonomy, data can be classified into two types. They are,
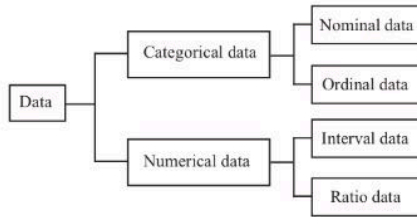
1. Categorical data

2. Numerical data.

Figure: Taxonomy of Data in Data Mining

1.   Categorical Data

Categorical data refers to the type of data which is divided into groups such as age group, sex etc. If a category of data carries a specific set of values without following any sequence then such categorical data can be referred as discrete data. The values used in these groups are considered as symbols even if they are numerical values.

Categorical data can be further categorized into nominal data and ordinal data.

(i)   Nominal Data

Nominal data can be viewed as objects which are measurements of codes allocated as labels. It can carry binomial or multinomial values. Binomial values carry two values such as True or False, yes or no etc., whereas multinomial values carry more than two values such as Red, Green, Blue, White.

(ii)   Ordinal Data

Ordinal data can be viewed as objects or events which are allocated with codes as labels representing ranks. An example of such data is Grade i.e., A, B or C. Where A grade refers to the score above 80, B refers to the score between 50 to 70 and C refers to the score below 50. Using such data, improvised classification models can be built.

2.   Numerical Data

Numerical data refers to the group of data carrying specific numerical values such as distance, age, salary etc. The values of this data can be in a particular sequence and can carry any number of values.

Numerical data can be further categorized into interval data and Ratio data.

(i)   Interval Data

Interval data carries values which are resulted from interval scales such as celsius scale. The result of this scale is the $(1/100)^{th}$ difference between melting point and boiling point of water at certain temperature.

(ii)   Ratio Data

Ratio data carries measurement values such as angle, length, energy, change etc.

2.2.2 Data Mining Functions

Q15. Discuss the Functions of Data Mining.

Answer :

The tasks or functions of data mining are classified into the following categories.

1. Prediction

2. Association

3. Clustering.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD. ――――――――――――――

1. Prediction

The term prediction refers to the action of predicting future. It involves various factors such as suggestions, expert advices, related information etc.

A term closely related to prediction is forecasting which is used as a synonym in most of the cases. The difference between prediction and forecasting is that the former is associated with experience and opinions whereas the latter is associated with data and model.

The most common type of tasks that comes under this category is classification

Classification

A large database has huge amount of raw data, which is analyzed and predicted to retrieve useful information and to make decisions. Classification is one of the methods used for data analysis. The user will analyze the data and classify it, based on the requirement. For example, if a person want to know the performance of the university, then he/she is required to classify the students database based on their performance as above average, average and below average students. If the classification shows that the number of students under "below average" category are more, then the university needs to improve.

2. Association

Associations are used to identify the variables present in the databases which are related based on certain condition. The technique of using associations is also called as association rule learning. However in retail field, it is called as market-basket analysis. In this field, bar code machines and point of scale systems are most commonly used for gathering data that helps in generating knowledge.

Association rule learning can be of two types. They are sequence mining and link analysis. The former one is used to analyse the relationship in a particular sequence so that associations are determined with respect to the time of their occurrence. The latter one is used to determine the relationships existing among objects.

3. Clustering

Clustering is a technique of combining a group of physical objects into classes of homogenous objects. Clustering of data objects into single class is equivalent to data compression. The class labels in this technique are not known. Different algorithms generate different clusters because these algorithms might consider different factors while generating the clusters. Therefore, an expert is required to select the type of clusters based on the requirements. Among all the available algorithms, k-means clustering is the most common one.

2.2.3 Data Mining Applications

Q16. What are the major application areas for data mining?

Answer :

The major application areas for data mining are as follows,

1. Customer Relationship Management

Data mining can be used to manage data such as sales, services, inquiries etc., which are related to customer relationship management. Data mining helps in performing the following operations in CRM.

❖ It determines the customers who interestingly participates in buying the goods/products.

❖ It helps in gaining customer retention by determining the causes of customer attrition.

❖ It determines the associations or relationships existing among goods and services with respect to the time.

❖ It determines the potential customers along with their needs to increase profit and improve relationship with the customers.

2. Banking

Data mining can be used in the field of banking in the following ways,

❖ It helps in making the loan application process automatic.

❖ It can capture the unauthorized transactions done over credit card or some other means of payment.

❖ It helps to forecast the flow of cash in various banking systems such as ATM, banks etc.

3. Retailing Industry

Data mining can be used in the field of retailing in the following ways,

❖ It can predict almost accurate inventory for a specific location by predicting the sales.

❖ It can use market-basket analysis to determine the products relationship thereby optimizing the sales.

❖ It can determine the degree of consumption for various types of products.

4. Manufacturing

Data mining can be used in the field of manufacturing in the following ways.

❖ It can make use of sensory data for predicting the errors in the machines.

❖ It can optimize the capabilities of manufacturing by determining the commonalities and anomalies.

❖ It can improve quality of goods by identifying novel patterns.

5. Brokerage and Trading

Data mining can be used in the field of brokerage and trading in the following ways,

❖ It can be used for predicting the time and reason behind the change in bond prices

❖ It can predict the impact of various issues on market

❖ It can determine fraudulent activities.

6. Computers

Data mining can be used in the field of computers in the following ways,

❖ It can determine the failures of disks and other devices

❖ It can capture and remove unnecessary e-mails.

❖ It can determine security needs and unsecure softwares.

7. Military

Data mining can be used in the field of military in the following ways.

❖ It can predict the cost required to move military equipments from one place to another.

❖ It can predict the consumption of resources.

❖ It can generate knowledge from data related to strategies and past experiences.

8. Travel industry

Data mining cars be used in the field of travelling in the following ways,

❖ It can predict the preferences of user with respect to different services such as types of seats, types of cars etc., to increase the revenue.

❖ It can also predict the demand with respect to various locations.

❖ It can determine potentially valuable customers.

2.2.4 Data Mining Techniques and Tools

Q17. Explain various classification techniques of data mining.

Answer :   *Model Paper-II, Q7(b)*

Following are various classification techniques of data mining.

1. Decision Trees

A decision tree is a tree structure consisting of following three types of nodes,

(i) Root Node

A node that do not has any incoming edge or zero or more outgoing edges.

(ii) Internal Node

A node that has only one incoming edge and two or more outgoing edges.

(iii) Leaf Node (or Terminal Node)

A node has only one incoming edge and no outgoing edge. In this, each non-leaf node (i.e., root node and internal node) represents the test on an attribute, each leaf nodes represents the class label and branches represent the outcome of the test.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD. ────────────────────

Consider the following example in order to understand the classification by decision tree.
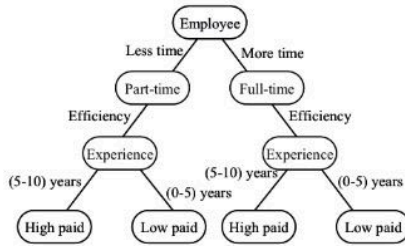

Figure: Decision Tree

In the above decision tree, the attribute "employee" is treated as a root node. This root node consists of two child nodes i.e., part-time and full-time. These child nodes becomes the parent node for "experience" node which inturn has two child nodes i.e., high paid and low paid. Here, classifying an high paid and low paid employee is done based on following rules,

1.   Whether the employee is working as a part-time employee or a full-time employee.

2.   Whether the experience of the employee is between 0-5 or 5-10 years.

   (i)   If the experience is between 0-5 years then he/she is considered as a low paid employee.

   (ii) If the experience is between 5-10 years then he/she is classified as high paid employee.

2.   Statistical Analysis

Statistical Analysis is one of the most easiest and conventional data mining technologies than the other technologies. One of the static model technique is Regression Analysis which is used for developing linear models. These models are used to determine the predictive data points. However, traditional statistical method can generate incorrect results in the following two situations:

   (i)   If the description of data is incorrectly provided by linear model.

   (ii) If the set of data items consists of many outliers.

Data mining employs the statistical technology which has the ability of handling non-linear, multiple outliers and non-numerical data present within the data warehouse environment.

3.   Neural Networks

A Neural Network is a collection of input and output units. The connection between these units have associated weights. During the learning stage, the weights of desired connections are adjusted so that the correct class labels can be predicted. This network describes the way a human brain works in recognizing the patterns. It does this by designing mathematical structures that have the ability to learn.

4.   Genetic Algorithm

Genetic Algorithm is a machine learning technique. Which works by creating a 'population' containing rules which are randomly generated. Each of the rules are expressed as a string of bits.

Consider a sample containing two attributes A_1 and A_2 and two classes C_1 and C_2. Suppose these attributes and classes belong to a particular training set. If there is a rule that "If A_1 and Not A_2 then C_1", then the corresponding string of bits is 101. Here, the attributes A_1 and A_2 are represented by the left most two bits and the classes are represented by the right most bit. Similarly, for the rule "If A_1 and Not A_2 then C_2", the string of bits is 100.

5.   Rough Sets

Rough sets are used when the classification is to be performed on the inaccurate or noisy data to find out the structural relationships among them. This approach can be applied only on the discrete-valued attributes. If the attributes have continuous values then they are transformed to discrete values before using this approach.

The rough set approach creates equivalent classes for the given training data. An equivalent class contains data tuples that cannot be distinguished from each other.

Q18. Write short notes on the following,

    (i) Cluster Analysis

    (ii) Association Rule Mining.

Answer :

(i) Cluster Analysis

The process of collecting homogeneous data objects within the same group called cluster and heterogeneous to the objects in other groups (clusters) is called as cluster analysis.

Clustering is a technique of combining a group of physical objects into classes of homogeneous objects. Clustering of data objects into single class is equivalent to data compression.

Various fields such as data analysis, market research, image processing, pattern recognition, business, statistics and biology uses the concept of cluster analysis.

On the basis of 'similarity', huge data sets are partitioned into groups. This process is known as 'data segmentation' which is equivalent to clustering. Outlier detection applications use clustering, these include the detection of credit card fraud and the monitoring of criminal activities in electronic commerce.

Cluster analysis is a function of data mining which may be used as a unique tool to attain knowledge about the distribution of data, to note the features of every cluster and to carry out further analysis by paying attention on a particular group of clusters. Cluster analysis can also be used as a preprocessing step for characterization, attribute subset selection and classification algorithms. Clustering can also be used to help in classifying documents on web for detecting information.

There is a strong development in clustering of data. Statistics, spatial database technology, marketing, biology, data mining and machine learning fields are contributing a lot towards the research of cluster analysis.

Cluster analysis as a branch of statistics focuses mainly on distance-based cluster analysis. Tools of cluster analysis that are based on k-means, k-medoids and various methods have also been built into many statistical analysis software packages like S-plus, SPSS and SAS.

(ii) Association Rule Mining

Association Rules are specified over both categorical and numerical attributes. These comprise of two statistical measures-'utility' and 'certainty'. These measures represent the usefulness and confidence of the extracted association rule. There are two threshold conditions set by experts, which must be satisfied in order to generate relevant association rules.

Association rules are expressed as $A_1 \Rightarrow A_2$ where, A_1 and A_2 represents conjuctions operation performed on certain conditions. These conditions can either be $m_i = c_i$ or $m_i \in [low_i, up_i]$. The variable C_i represents either categoric or numerical value and low_i, up_i represents the interval ranges of numerical attribute. The level of support for rule $A_1 \Rightarrow A_2$, is equal to the conjunction support level i.e.. $A_1 \cap A_2$. The rule $A_1 \Rightarrow A_2$ means that, if a transaction contains A_1, then it should also contain A_2.

Association rule can be mined by considering the following steps,

(a) Search for Item Sets that Occur most Frequently

Frequent Item Sets are defined as a set of items, that occur as frequently as, predefined minimum support threshold items.

(b) Produce Strong Association Rule from the Resultant Item Set

Association Rule, must satisfy both Minimum Support Threshold, as well as Minimum Confidence threshold conditions.

Q19. Discuss in brief about the following algorithms,

    (i) K-means Clustering

    (ii) Apriori.

Answer :

(i) K-means Clustering

In this algorithm, the mean value of the objects in the cluster represents a cluster.

The k-means is an iterative clustering algorithm in which objects are moved among sets of clusters until the desired set is achieved. It is the most popular and commonly used method. The algorithm is built on the concept of user specified input parameter (k). A set of 'n' objects are divided into 'k' clusters by the algorithm. A high degree of similarity among elements in clusters is obtained, while a high degree of dissimilarity among elements in different clusters is achieved simultaneously. The cluster's centroid gives the measure of cluster's similarity.

Step1: Initially select 'k' objects randomly from D, as initial cluster centers.

Step2: Depending upon the distance between the object and the cluster mean, each remaining object is assigned to the cluster to which it is most similar or near.

Step3: Calculate new mean value of the object for each cluster.

Step4: The step (3) is repeated and the process iterates until the criterion function converges. The resulting 'k' clusters are compact and separate. k-means method, typically, uses the square-error criterion function, which is expressed as,

$$S = \sum_{a=1}^{k} \sum_{e \in C_a} |e - \text{Mean}_a|^2 \quad \text{and}$$

$$\text{Mean}_a = \frac{1}{\text{Mean}} \sum_{j=1}^{\text{Mean}} n_{aj}$$

Where,

S = The sum of the square error for all objects in the data set.

e = The point in space that represents the given object.

Mean_a = The mean of cluster C_a.

Example

Consider a triangular space containing a group of numbered objects. Let, $k = 2$ be the numbers of clusters desired by the user. The algorithm proceeds as follows.

Randomly choose two objects as initial cluster centers marked as ✿ in the figures with mean values as $m_1 = 2$ and $m_2 = 4$. And find the Euclidean distance between the mean and the objects to classify objects into two clusters as shown in figure (1).
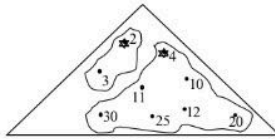

Figure (1): Initial Partitioning with $m_1 = 2$ and $m_2 = 4$

Now, evaluate new mean for the resulting clusters and again partition the objects is done on the basis of euclidean distance as shown below,

$$m_1 = \frac{2+3}{2} = 2.5$$

$$m_2 = \frac{4+10+11+12+20+25+30}{7} = 16.$$

$m_1 = 2.5$ and m_2

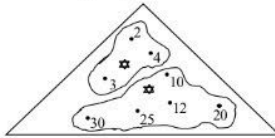= 16 and the resulting cluster is shown in figure (2).

Figure (2): New Clusters Formed with $m_1 = 2.5$ and $m_2 = 16$

The process iterates for the successive values of the mean, to enhance the partitioning of the clusters. At last there is no redistribution of objects in any cluster and the process terminates when the means become identical and converge. The following figures depict the iterative relocation.
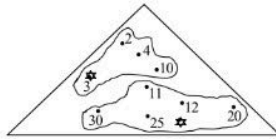
Figure (3): Clusters with $m_1 = 3$ and $m_2 = 18$
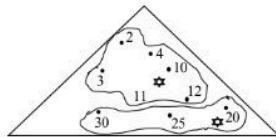
Figure (4): Clusters with $m_1 = 4.75$ and $m_2 = 19.6$
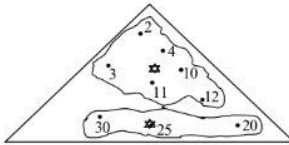
Figure (5): Final Clusters are Returned by the Clustering Process with $m_1 = 7$ and $m_2 = 25$

Thus, finally the correct partitioning of the numbered objects into two clusters is achieved as desired by the users.

The time complexity of k-means is $O(nkt)$, where 'k' is the number of cluster, 't' is the number of iterations and 'n' is the total number of objects. The k-means finds a local optimum. It does not work on categorical data because the mean must be defined on the attribute type. It can find only convex-shaped clusters. k-means also does not handle outliers well. Although, the k-means algorithm produces good results, it is not time-efficient and does not scale well.

(ii)   Apriori

Apriori Algorithm, is the most preferred algorithm, for mining single dimensional boolean association rule. This algorithm, exploits the prior knowledge associated with the characteristics of itemsets, which are occurring frequently. Apriori is based on iterative process, so as to find out itemsets present at (n+1) level by scanning n-level itemsets. The process is initiated by scanning the frequent itemsets at level-1 and combining every item count present at the same level. The only items that don't violate minimum support threshold conditions are accumulated. The resultant items are used to scan frequent items, present at the next successive level. The process is repeated, until a level, which doesn't contain any item set is reached.

Apriori Algorithm, makes use of Apriori property so as to enhance the efficiency of frequent itemsets generated level-by-level. Another benefit of Apriori property is that, it decrease the size of search space. This property states that, "Every subset of frequent itemset must also occur frequently".

Q20. List out various software tools of data mining.

Answer :   *Model Paper-III, Q7(b)*

Some of the software tools of data mining are,

1.   IBM Intelligent Miner

2.   SPSS Clementine

3.   Oracle Data Mining (ODM)

4.   Classification and Regression Tree (CART)

5.   Weka.

1.   IBM Intelligent Miner

IBM intelligent miner offers many data mining functions such as classification, association mining, predictive modeling, clustering, deviation detection, sequential pattern analysis and regression.

This tool also offers an application tool kit that contains statistical method, data visualization tools neural network algorithms and data preparation tools. It is highly scalable with mining algorithms and can be tightly coupled with IBM's DB2 database system.

2.   SPSS Clementine

SPSS clementine offers an integrated development environment for developers and end users to perform data mining. It performs various functions including prediction, clustering, association mining and classification. It is also equipped with object oriented module interface through which users can add their utilities to visual programing environment.

3.   Oracle Data Mining (ODM)

Oracle Data Mining (ODM) tool is an add on feature of Oracle 10g Enterprise Edition database. This tool performs various data mining functions such as classification, prediction, sequence similarity search and analysis, association mining and clustering.

4.   Classification and Regression Trees (CART)

Classification and Regression Trees (CART) is a data mining tool from Salford systems. This tool generates regression trees for prediction and decision trees for classification. It improves the accuracy of data mining by employing boosting mechanism.

5.   Weka

Weka is an open-source data mining tool developed on Java platform in the university of Waikato, Newzealand. It is capable to perform various data mining functions such as regression, classification, association mining and data preprocessing.

2.2.5 Text Mining

Q21. Define text mining. List the application areas of text mining.

Answer :

Text Mining

Text Mining is defined as a process of extracting quality oriented information for text (document) databases. The main purpose of this mining is to process unstructured information and to extract meaningful numeric indices from the database, so as to make the information accessible to different data mining algorithms. Text mining is an important part of data mining process because such mining enables the user to make comparison among several documents, provide priority to essential documents or identify the procedure of several documents.

The functioning of text mining is identical to data mining that considers the input in the text format only. The files can be MS-Word documents, PDFs, XML files etc. It is considered as effective in the fields where text data is generated in a larger amount. Examples of such fields are court orders, academics, medicine, marketing etc.

Application Areas of Text Mining

Some of the major application areas of text mining are as follows,

❖   Key phrases and various relationships are extracted from text databases with use of pattern matching approach. This is called information extraction.

❖   Documents which might be useful to the users are predicted based on the views and profile of users. This is called topic tracking.

❖   Descriptions are provided for the documents to save time. This is called summarization.

❖   Documents are classified based an certain predefined categories. This is called categorization.

❖   Interrelated documents are grouped together which is called clustering.

❖   Interrelated documents are linked together so that they can be easily searched. This is called concept linking.

❖   Knowledge based pattern matching approach is adopted in order to solve queries.

Q22. Discuss the application categories of text mining.

Answer :

The following are some of the application categories of text mining.

1. **Marketing Applications**

    Text mining can be used to assess the data generated from the call centers of the organizations in order to improve cross -selling. This is because the generated text contains customer feedback regarding the products of the organization. This information can be analyzed to predict customer perceptions. the same type of information can be extracted by blogs of websites, e-mails and discussion boards. Real time application of text mining in marketing is the one experienced by Ghani.

2. **Security Applications**

    An example application of text mining in security domain is ECHELON system which is capable of tracking information transferred through telephone calls, faxes, e-mails etc.

    EUROPOL is another application which is designed to capture the criminal activities by storing and assessing massive amount of sources that generate data in both structured and unstructured form.

    Text mining can also be applied to the statements recorded from criminals to capture true statements.

3. **Biomedical Applications**

    The reports or data generated from various tests such as microarray analysis, SAGE etc., are in rich textual format. This opens the door for text mining to analyze the large amount of data generated from these experiments. The analysis is performed in order to carryout validation and interpretation.

    Shatkay suggested a system to determine the location of proteins based on text mining systems. This system captures and integrates the text sources by making use of sequence based features. This system is considered as effective over the previously generated outputs.

    Another example application of text mining in biomedical industry is the system developed by chun. This system was capable of generating relationships between disease and gene. This system uses a dictionary carrying names of diseares and genes and then matches these names. The false positives are eliminated using NER (Named Entity Recognition) approach.

4. **Academic Applications**

    One application of text mining in academic field is the collaboration among universities present in Manchester and the National centre for Text mining. They decided on generating customized tools, facilities and suggestions on academics using text mining approach.

Q23. Describe the three-step text mining process.

Answer :

The three major steps involved in text mining process are,

1. Establishing the Corpus

2. Developing the Term-Document matrix

3. Extracting the Knowledge

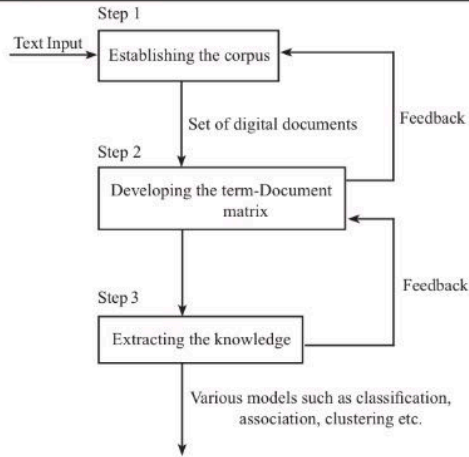The typical process of text mining is illustrated in the figure below,

Figure: Task-mining Process

1. Establishing the Corpus

In this step, the extracted text documents such as word documents, HTML files, XML files etc., including voice recordings are added to the text collection. These documents are arranged and transferred into a specific format so that it can be further processed by the computer. The text documents in the digitized format are forwarded to the second step as its input.

2. Developing the Term-document Matrix

In this step, a term-document matrix is developed from the digitized documents provided as input to this step. This matrix contains a set of rows and columns where rows represent documents and columns represent the terms as shown below.

	Risk of Investment	Software Engineering	SAP	Development
DOC1	1		1	
DOC2		2		
DOC3	1		1	2
DOC4		3		

The cells between rows and columns of the matrix carry indices. The values of indices are nothing but the frequency of the terms occuring in the documents. The common terms such as articles, auxillary verbs are excluded and the terms considered in the matrix are called stop terms. Synonyms can also be included to make the search appropriate.

To perform proper filteration in indices, stemming method can be used. This method considers all the forms of a verb as a single word thereby minimizing the number of words. An example of stemming is word pick which considers all its verb forms i.e., picked, picking.

3. Extracting the Knowledge

In this step, the terms-document matrix generated from step2 is integrated with knowledge extraction methods such as classification, association, clustering etc., to generate knowledge from the TDM.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

2.2.6 Web Mining

Q24. Define web mining. Explain the three main areas of web mining.

Answer :

Web Mining

Web mining refers to the extraction of knowledge from world wide web (www). The web is considered as a network of networks where massive amount of data is stored, retrieved and searched. It is difficult to perform knowledge discovery on the web because of its huge size, complexity, dynamic and diverse nature.

Areas of Web Mining

The three main areas of web mining are,

1. Web content mining

2. Web structure mining

3. Web usage mining

1. Web Content Mining

Web content mining is a process of extracting relevant information from web contents. Basically, web content comprises not only textual information but also graphical information, real-time information and hyperlinks. The textual information in web content data is a combination of unstructured (free text), semi-structured (HTML pages) and highly structured (database generated HTML pages) data. However, web content data is unstructured due to which text mining techniques can be used for performing web content mining.

2. Web Structure Mining

Web structure mining is the process of generating summary of webpages of websites. Web structuring checks link structure i.e., hyperlinks among various websites and classifies the webpages on the basis of hyperlinks found. It discovers the relationship of webpages and the pages to which the links are found. The relationship is determined on the basis of synonyms or similar contents found on the webpages.

It also determines the network in a certain domain. This determination makes the process of querying more easier and efficient.

3. Web Usage Mining

The Web usage mining simply refers to the process of searching the Weblog records in order to determine the procedures in which users can access the Web pages.

If the Weblog records database are examined carefully, it

(a) Determines all the possible e-commerce customers

(b) Efficiently delivers the qualitative Internet Information services to the user

(c) Enhances the performance of the Web server.

The Web server will record a new Weblog entry every time the Web page is accessed. This web log entry contains the information about the address of the web page (i.e., URL), IP address of the requester, and a timestamp.

The most frequently referred Websites and Web-based e-commerce servers need to maintain a large Weblog database which consists of information about millions of Weblog records. This Weblog database can be accessed with the help of Weblog mining techniques.

SHORT QUESTIONS AND ANSWERS

Q1. Define data warehouse.

Answer :   *Model Paper-II, Q2*

Data warehouse is a form of storage system (database) where large volume of data is stored in such a way that retrieving desirable information from the system is very easy and reliable. Data warehouse is stored in different location so that it doesn't collide with transactional database system which stores day-to-day information and answers the queries that are prerecorded in the database. Data warehouse system on the other hand provides solutions to sophisticated queries, which involves many computations to be performed at finer-level of granularity.

Q2. Explain why we need to separate data warehouse.

Answer :

Data warehouse is kept separate from operational system because of the following reasons.

(i)   Data warehouses are optimized to execute select-type queries whereas, operational databases are optimized to execute insert and update type queries.

(ii) Data warehouse schemes are simplified and denormalized whereas, the schemes of operational systems are large and complex.

(iii) Data warehouse stores historic information whereas, operational system stores only day-to-day information.

(iv) Data in data warehouse are non-volatile (i.e., do not change) whereas, data in operational system is highly volatile.

Q3. What are the advantages of data warehouse?

Answer :   *Model Paper-I, Q2*

The advantages of data warehouse are as follows,

❖   Data warehouse is capable of storing and consolidating past information.

❖   It provides support for sophisticated multidimensional queries.

❖   It increases the performance of integrated database system as data from heterogeneous sources are extracted, preprocessed, cleaned, transformed into one unified data store.

❖   It doesn't use query driven approach (as it requires difficult filtering and integration techniques) instead uses update-driven approach where in the data collected from various sources are consolidated and stored in warehouse for performing explicit data analysis.

❖   It has the ability of supporting management decision activities like managing relationship with potential customers, managing the cost of important assets, evaluating operation and searching for sources using which profit can be attained.

❖   It is capable of understanding the current business trends and making better forecasting decisions.

Q4. What are data marts?

Answer :

Data marts contain versatile set of data which is selected from enterprise wide data. The data is selected based on the requirement of particular user group. It represents data from a single process of business. The difference between enterprise warehouse and data mart is that, the former is cross-functional in scope, but the latter is limited to particular selected subjects. The data in data marts are not in detailed format but in summarized and simplified form.

Data marts are divided into two-types. They are,

(a)   Independent data mart

(b)   Dependent data mart.

Q5. Write short notes on data mart bus architecture and hub-and-spoke architecture.

Answer :

Data Mart Bus Architecture

This architecture is an extension of independent data marts architecture. In this architecture, the data marts are connected together through a middleware. Using such an architecture data consistency can be achieved. It can process complex queries but does not fulfil the performance requirements.
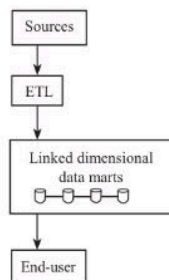
Figure: Architecture of Data Mart Bus

Hub-and-Spoke Architecture

The main objective of this architecture is to achieve scalability and maintainability within the infrastructure. To achieve this, different subject areas are considered one after the other in an iterative way. It carries a centralized warehouse and multiple data marts which are interdependent. The advantage of this architecture is its simplicity and customization support for interfaces and reports. However, the drawbacks of this architecture includes redundancy and latency of data.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.
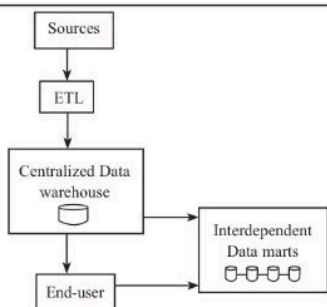
Figure: Architecture of Hub-and-spoke Dataware Housing

Q6. Define OLAP.

Answer :

OLAP comprises of set of standards that are responsible for providing dimensional structure for supporting decision system. The main purpose of OLAP is to perform data analysis and to access data on-line. It provides user-friendly interface for evaluating data interactively. OLAP are tools which are dependent on multidimensional database conception. It enables a highly developed users to evaluate the data using complicated and sophisticated views. They assume that data is arranged in multidimensional model, which is backed-up by relational databases. OLAP system consists of more complicated query outcomes than transactional database system. This analytical processing is performed on data warehouses which involve analysis of actual data.

The main objective of OLAP system is to assist ad hoc querying required to support decision support system. They are a type of software technology that allows system analysts, manager to understand the data using fast interactive technologies. In this, access is made from a large spectrum of potential view of information, which is generated from raw data. Such access indicate the actual dimensionality of an organization from users perspective.

Q7. What is data cleaning?

Answer :

When data is collected from real-world then there are chances for the data to be inconsistent, incomplete and noisy. Data cleaning is a process of removing unnecessary and inconsistent data from the databases. When data is extracted from heterogeneous data sources, then there are chances that same information may be in different formats and metrics. The main purpose of data cleaning is to improve the quality of data by filling the missing values, reconfiguring the data to make sure that data is in consistent format.

Q8. Write a short note on data discretization.

Answer :

Class intervals are constructed by partitioning the discrete range of attribute values in order to decrease the total number of attribute value sets. This is done using a technique known as data discretization. The main objective of this approach is to substitute continuous values with limited number of interval description. The advantage of this substitution is that it makes the actual data to be easily interpreted which results in precise and simple representation of mining results.

Q9. List the major issues in data mining.

Answer :

There are many implementation issues related to data mining which arise when databases are incomplete, noisy and contain of missing values. There are other issues as well that arise due to insufficient and inappropriate information present in the databases.

The issues in data mining are classified as,

(i) Limited and irrelevant information

(ii) Noisy and missing data

(iii) Human interaction and prior knowledge

(iv) Large data sets and high dimensionality

(v) Uncertainty

(vi) Dynamic updates.

Q10. What is meant by classification? What are applications of classification model?

Answer :　　　　　　　　　　　　　　*Model Paper-III, Q2*

Classification

A large database has huge amount of raw data, which is analyzed and predicted to retrieve useful information and to make decisions. Classification is one of the methods used for data analysis. We analyze the data and classify it, based on our requirement. For example, if we want to know the performance of the university, we classify the students database based on their performance as above average, average and below average students. If the classification shows that the number of students under "below average" category are more, then the university needs to improve.

Applications of Classification Model

The classification model can be used in,

(i) Modeling of business

(ii) Segmenting of customer

(iii) Analysing of credit.

INTERNAL ASSESSMENT

I. **Multiple Choice**

1. _____ is a repository for long term storage of data from multiple sources and enables management in decision making.   [   ]

    (a) Relational database

    (b) Data warehouse

    (c) Transactional database

    (d) Object-relational database

2. _____ is used for discovering interesting patterns from large amount of data.   [   ]

    (a) Data mining

    (b) Data integration

    (c) Data transformation

    (d) KDD

3. _____ mining tasks perform inference on the current data in order to make predictions.   [   ]

    (a) Predictive

    (b) Descriptive

    (c) Subjective

    (d) Informative

4. _____ mining tasks characterize the general properties of the data in the databases.   [   ]

    (a) Descriptive

    (b) Predictive

    (c) Meta data

    (d) Data

5. The compressed form of input data is called as _____.   [   ]

    (a) Binary tree

    (b) Fp-tree

    (c) Prefix tree

    (d) Subtree

6. The _____ of an itemset identifies how frequenty a rule is applied to the provided dataset.   [   ]

    (a) Count

    (b) Support

    (c) Confidence

    (d) Rule

7. The process of mining the association rule can be done by employing _____.   [   ]

   (a) $F_{n-1} \times F_1$ method

   (b) $F_{n-1} \times F_{n-1}$ method

   (c) Association rule

   (d) Brute-force approach

8. _____ is a type of data warehouse.   [   ]

   (a) Data mart

   (b) ODS

   (c) EDW

   (d) All the above

9. Tier-3 of data warehouse architecture represents _____.   [   ]

   (a) Front-end client layer

   (b) Application server

   (c) Warehouse server

   (d) None of the above

10. _____ integrates business processes by dividing large application into smaller modules or services.

                                                                                                       [   ]

    (a) SOA

    (b) EAI

    (c) EII

    (d) ETL

II.   Fill in the Blanks

1. Data warehouse contains _____ data.

2. _____ is an interface between the data warehouse and decision support system.

3. _____ is a process of extraction, transformation and loading.

4. Smooting of noisy data is done through _____.

5. Single dimensional association rule is mined using _____ algorithm.

6. Classification is one of the method _____.

7. _____ of a decision tree has only one incoming edge and no outgoing.

8. ODS stands for _____.

9. Tier-1 of data warehouse architecture represents _____.

10. The Inmon model is _____ approach.

KEY

I.   Multiple Choice

1. (b)

2. (a)

3. (a)

4. (a)

5. (b)

6. (b)

7. (d)

8. (d)

9. (c)

10. (a)

II.   Fill in the Blanks

1. Non-volatile

2. Metadata repository

3. ETL

4. Data cleaning

5. Apriori

6. Data analysis

7. Leaf node

8. Operational data stores

9. Front-end client layer

10. EDW

III.   Very Short Question and Answers

Q1.   What types of data is tored in ODS?

Answer :

The operational data stores carry the most recently updated information regarding the customers. This information is updated throughout the life cycle of a business operation.

Q2.   Describe tier-2 of data warehousing architecture.

Answer :

This tier represents the application server in which data acquisition software resides. This software is responsible for extracting the data from various sources and storing it into the warehouse after transformation.

Q3.   What is real-time data warehouseing?

Answer :

Real-time data warehousing is a type of data warehousing in which the data is processed and loaded instantly when it becomes available. For this reason, it is also called as active data warehousing.

Q4.   Define nominal data.

Answer :

Nominal data can be viewed as objects which are measurements of codes allocated as labels. It can carry binomial or multinomial values. Binomial values carry two values such as True or False, yes or no etc., whereas multinomial values carry more than two values such as Red, Green, Blue, White.

Q5.   Define clustering.

Answer :

Clustering is a technique of combining a group of physical objects into classes of homogenous objects. Clustering of data objects into single class is equivalent to data compression. The class labels in this technique are not known.

Business Performance Management (BPM)

After studying this unit, one would be able to understand,

* Concept of Business Performance Measurement.
* Differences between BPM and BI.
* Various BPM Processes.
* Performance Measurement.
* BPM Applications.
* Performance Dash Boards.

INTRODUCTION

Business process management can be defined as a sum total of tools techniques and matrics used to measure and manage the organizational performance. It is also called as corporate performance management or enterprise performance management. In comparison to business intelligence, business performance management is wider. Some researches state that BPM is an outgrowth if B.I.

Strategic gaps refer to the differences (gaps) between an organization strategies and actual strategies achieved by organization. The reason for such gaps could be lack of communication resource or time crunch. To understand, analyze and adapt the enterprise according to the positive or negative responses of the customers, one can employ the harrahs closed loop marketing model.

Performance measurement system is also an integral part of business performance management. The different tools of PMS includes activity based costing, six sigma, balanced score card and so on.

The focus of six sigma is to produce defect free good or services along with continuous improvement whereas balance score card has been designed and developed to balance the organizations strategic elements such as customers internal processes, finance, learning and development.

Performance dash boards provides visual display of vital information on a single screen to the top managers and facilitate in the decision making process.

3.1 BPM – DEFINITION AND COMPONENTS

Q1.   Define Business Performance Management (BPM)? State the features and components of BPM.

Answer :   *Model Paper-I, Q8(a)*

Business Performance Management (BPM)

BPM may be defined as "Different business tools, techniques, methodologies and metrics employed by an organization to measure, moniter and manage the performance of the business enterprise". There is always a confusion between the term Performance Management and Business Performance Management (BPM). However both are interlinked with each other and in most cases used as synonyms.

According Dr. T.V Rao, "performance management involves, thinking through various, facts of performance, identifying critical dimensions performance, planning, reviewing, developing and enhancing performance and related competencies".

According to Reward Management Associates "performance management may be defined as managerial process which consist of planning performance, managing performance through observation and feedback, appraising performance and rewarding performance".

In simple words we any define performance management as "a continuous process of identifying, measuring and developing the performance levels of individuals and aligning it with the strategic goals of the organization.

Features of BPM

Some key features related to business performance management are,

(a)   BPM is known by other names such as Corporate Performance Management (CPM), Enterprise Performance Management (EPM), Strategic Performance Management (SPM).

(b)   The term CPM was coined by the market analytics firm Gartner.com. Whereas the term BPM was first used by the BPM standards group.

(c)   BPM employs a wide range of tools, techniques and metrics to assess the actual performance of the business organization.

Components of BPM

According to business analytics researcher colbert (2009) following are the major components of BPM,

(i)   A key component of BPM is a set of closely linked management and analytical processes. These processes are fully integrated with technology. The primary focus area of these close looped management processes is to address and assess the financial and operational activities of the organization.

(ii)   BPM components include all the business tools and techniques which are used for defining, formulating and measuring the strategic goals of the organization. These tools and techniques are used for not only formulating strategic goals, but also for measuring the actual organizational performance with the previously set goals.

(iii)   Other important components of BPM includes tools and techniques of financial and operational planning, metrics employed for the measurement of key performance indications and so on.

Thus, it may be stated that business organizations operate in a highly competitive environment. Thus, they need to adapt, adjust and align themselves on a continuous basis as per the changing business environment. Researcher Axson (2017) had stated that "An organization's performance management processes are the principle mechanism for assessing the impact of change and turning the business in order to survive and prosper.

3.2 BPM VS BI

Q2.   Compare and contrast BPM (Business Performance Management) and BI (Business Intelligence).

Answer :

The following are the key differences between BPM and BI,

	Area	B.P.M	B.I
1.	Definition/ Meaning	BPM refers to the different business tools, methods and metrics employed by an organization to measure, moniter and manage business performance.	BI refers to the art and science of analysing any data or information, with a specific business objective.
2.	Scope	The scope of BPM is wider. Impact BPM can be considered as an outgrowth of BI.	The scope of BI is comparatively narrower.

3.	Tools used	BPM employs multiple tools, methods and metrics, which includes all the BI tools as well.	BI employs a wide variety of tools and techniques such as OLP adhoc querying, dashboards, score-cards etc.
4.	Processes involved	It involves multiple processes including BI+, planning, cycle of plan, monitoring and analysing of business performance.	BI processes are limited only upto analysing of business data.
5.	Nature	BPM is highly flexible in nature and is actively integrated with the ongoing organizational projects. It would be aligned as per the changing business needs.	BI is static in nature. Its primary focus is upon measuring the performance of already completed business projects.

A detailed compare and contrast between BPM and BI would reveal that there exists many common denominators between the two. For example,

(i)   BI is a major element of BPM.

(ii)   The vendors who promote BPM are the same firms which sell BI tools and products.

(iii) Over the last few years, BI had evolved substantially. This had further blurred the difference between BPM and BI.

Although BI metrics and softwares are an integral part of BPM solutions, BPM has an impact over the organization wide strategy.

Researchers Tucker and dimon (2009) had stated that "BPM is not a one-off project or departmentally focussed. Instead BPM is an ongoing set of processes, that if done correctly, impacts an organization from top to boottom. If helps users take action in pursuit of their 'common cause' achieving performance targets, executing company strategy and delivering value to the stakeholders".

3.3 SUMMARY OF BPM PROCESS

Q3.   Discuss in detail major BPM processes.

<p style="text-align:center">OR</p>

Explain in detail about BPM cycle.

Answer :   *Model Paper-II, Q8(a)*

The Business Process Management (BPM) is not an individual tool or methodology. It is a set of repetative processes which are dynamic in nature and needs to be carried out on a regular basis, adjusting the various processes, as per the organizational requirements. The following figure shows the BPM process or cycle,
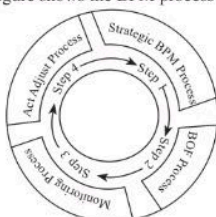
Figure: BPM Cycle

Step-1: Strategic BPM Process

This is the first step of the BPM cycle under which the organizations initiates a detailed analysis of the strategic gap that exists in the organization and formulates a detailed strategic planning. It includes the following steps.

(a)   The firm conducts a current situation analysis to determine its current position.

(b)   A SWOT (Strength, Weakness, Opportunities, Threats) analysis is conducted.

(c)   The organizations Critical Success Factors (CFS's) are identified.

(d)   The creation of strategic vision and the development of the business level strategy.

(e)   The organizations strategic goals and objectives are identified.

Step-2: BOF Process

Under the Budgeting, Operations and Financial (BOF) process, the strategic goals and objectives of the organization are converted into well defined tactical initiative at the sub process levels.

Operational planning comprises of tactics and initiatives which lead to the desired organizational level. A good operational plan could be either centered around the budget or tactic as per the organizational requirements.

Financial planning and budgeting play a vital role in the overall success of the organization, because the resources of score and needs to be allocated prudently and to extract maximum benefit out of them. If the focus is upon the tactic, then the organization may prepare a tactic based financial budge.

Step-3: Monitoring Process

Under this step of the BPM cycle, a good and effective performance monitoring mechanism is set in place to ensure that all organizational processes are working towards the desired organizational mission/goal. The various tools techniques and models of performance monitoring clearly showcases the deviations from the desired performance.

A diagnostic control system is a cybernetic system, that is, it intakes inputs, process the information and provide outputs. This output is then compared with pre-established standard performance benchmarks and feedback is provided stating whether there had been any variance from the expected performance.

All organization, irrespective whether they employ BPM system or not, would use some type of diagnostic control system. The key elements of this system are,

(a) Inputs

In the form of actual performance.

(b) Monitoring Elements

This include any type of tool, technique, metric or measure used by the organization for processing the information. Example: Balanced Score Card, Dashboards, Human Resource System, Financial management system etc.

(c) Output

This element comprises of any type of end result produced by the monitoring system.

Step-4: Act Adjust Process

The final step of the BPM cycle involves measuring the business performance by using a variety of tools, techniques, metrics and models and take corrective action (if needed).

For remaining answer refer Unit-IV, Page No. 3.6, Q.No. 6, Topic: Harrah's Closed Lool Marketing Model.

Q4.   State the basic tasks in the strategic planning process.

Answer :

The concept of strategy has a wider meaning and includes concepts such as strategic vision and strategic focus. Strategic planning answers the key question, where does the organization desires to see itself in the future? The corporate level strategy planning is an integral part of BPM.

When organisations are done with corporate planning they move ahead for business unit level planning which originates from the corporate plan. The business unit level strategic planning involves six steps as follows,

1.   Deciding on the Business Mission

As SBUs are operating in different market conditions, business mission needs to be established by considering both the overall corporate mission and objectives of the firm. Business mission must represent its motive of existence into business and about its role.

2.   Performing SWOT Analysis

SWOT analysis is conducted by the firm to evaluate strength, weakness, opportunity and threat of each and every business unit. Strengths and weakness analysis is done to analyse the internal strengths of the firm. Whereas, opportunity and threat analysis is done to analyse the external environment.

(a)   External Environment Analysis (Opportunities and Threats)

Both macro and micro environmental factors are analysed and monitored in the external analysis. Through such analysis, both potential opportunities and threats of the firm can be identified so as to optimally exploit the opportunities and to overcome threats. Market opportunity analysis is applied in determining the market attractiveness and probability of success of the opportunity.

(b) Internal Environment Analysis (Strengths and Weaknesses)

Internal evaluation is performed to be aware about the resources, behaviour, strengths, weaknesses, synergistic effects and distinctive competencies. It is an evaluation of the internal capability of the firm which can be optimally utilized for the exploitation of existing opportunities and for opposing the external threats of an environment.

3. Formulation and Implementation of Business Goals

The step ahead of the SWOT analysis involves the formulation of realisable and measurable goals for the business. Such goals are used to explain the objectives of business related to its marketing expenditure for a particular period of time. Achievement of a desired market share, profit, sales and level of reputation are some of the business goals.

4. Formulation of Business Strategy

The long-term goal directed actions are usually referred to as a "strategy". An appropriate strategy is selected by considering the strengths and goals formulated for the business unit. Goals indicate what is to be achieved whereas, strategy represents the courses of action taken to achieve these goals.

According to Michael Porter, Firms can follow three different generic strategies for the accomplishment of organisational goals. They are cost strategy, differentiation strategy and the focus strategy.

5. Formulation and Implementation of Program

After the business unit planning, the marketing manager needs to prepare comprehensive supporting programs. These programs need to be functional plans that are helpful in the implementation of strategies. Marketing managers must prepare a marketing plan which involves cost estimates, allocation of budget and investments related to a specific program. When program is implemented, it specifies the structures, responsibilities and role of every member of an organisation.

6. Feedback and Control

The final step in strategic planning process is to assess and analyse the entire process at different points of time. Feedback is useful for the determination of the market response towards the effectiveness of marketing strategy. The entire planning process can be controlled by several methods such as, cost control, performance control and adaptability control.

Q5. What is 'Strategy Gap'? What are the reasons/sources of strategy gap?

Answer :

Strategy gap refers to the difference between the organizations strategic plan and the actual execution of that plan. Thus, this difference between the strategic plan and the actually accomplished strategy refers to strategy gap. For example, if the strategic plan of an enterprise was to boost the sales by 100% in the next one year and if the sales data of the next years Shows an increase of only 30%, then the 70% of unachieved sales target is the strategic gap. The top management need to analyze the reasons for this huge strategic gap and address the same to eliminate or reduce it. The most common reasons for strategy gap are,

(a) Lack of Communication

The most common reason for strategy gap is lack of communication between the top management and the organizational workforce, with regard to the organizations strategy. According to the researching firm Plladium Group, less than 10% of the workforce is made aware of the strategic plan. Farther the strategic plan which is communicated generally lacks clarity.

(b) Sync between Rewards and Incentives

It is a major reason for strategic gap. According to research Norton (2007) 70% of organizations do not align the incentives of the middle level management to the success of the strategic plan. Many firms wrongly link pay to short term performance.

(c) Resource Crunch

Any grand strategic plan need equally strong funds and resources both in terms of finance and non-financial resources. According to researcher Norton, any good strategic plan, which is not funded sufficiently is doomed to fail. Statistical information had concluded that only 30% of the organizations sufficiently finance their strategic plans.

(d) Time Resource

Researcher Norton had proved that 90% of the top management spend less then 1 hour per day to discuss and focus upon the organization's strategic plan, this inturn would lead to strategy gap. Time is a key component/resource that need to be spent on a regular basis upon the organization's strategic plan.

Q6.   Describe the steps in Harrah's closed-loop marketing model.

Answer :

Harrah's closed loop marketing model was presented by Watson and Volonino in the journal "Harnessing customer information for strategic advantage, technical challenges and business solution" (2001). The main aim of this model is to analyze the customer responses and adjust the business offering accordingly to run the business successfully.

Research studies had established that the probability of failure of any business is between 60 percent to 80 percent. For example, the probability of failure for an IT business is 70% whereas, it is 90% in case of pharmaceutical products.

The reasons for the failure of the business may be the result of no proper competitor analysis, insufficient demand analysis, changing customer tasks and preferences and so on.

Harrah's Closed Loop Marketing System Model

As depicted in the figure below the Harrah's model comprises of five sequential steps,
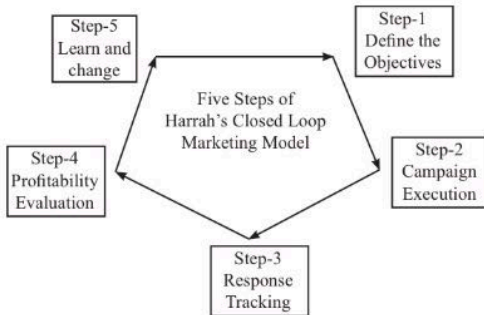
Step-1: Define the Objectives

It is the first step of the Harrah's closed loop marketing model. Under this step, the quantifiable objectives of the marketing campaign are defined or procedures are tested in the form of excepted values utilizing the customers in the test group verses the customers in the control group.

Step-2: Campaign Execution

Under this step the campaign is executed. The campaign is designed and executed in detail such that the optimum mix of offer and message is provided at the right time. The customers are selected after a careful review.

Step-3: Response Tracking

The next step involves tracking of the customers response. Other metrics are also employed to gauge the type of response provided by the customer, what type of incentive is generating a positive response and in how much monetary terms, profitability per customer, the quantity and at what price the customer had purchased the product and so on. This step involves a detailed tracking of the consumers response.

Step-4: Profitability Evaluation

Under this step the marketing campaign is evaluated in terms of the net profitability generated by the campaign and a comparative analysis is also conducted in relation to the response generated by other similar marketing campaign.

Step-5: Learn and Change

Under this final step of Harrah's closed loop marketing model, the information and knowledge gathered from the marketing campaign is analyzed, lessons learned and changes made accordingly, to boost and enhance the profitability and sales of the business organization.

Q7.   What are the findings of the Saxon groups research on performance management practices of an average company?

Answer :

The Saxon Group Consulting Firm (SGCF) is a consulting firm headed by David Axson. He is a global strategic advisor who is a specialist at measuring the performance of organizations, based upon various benchmarking parameters, Mr David had personally participated in the benchmarking programs of more than 300 or more enterprises worth more than $ 500 million. The SGCF had conducted performance surveys with more than 1000 top level executives, with an aim to gauge the organizational performance. These top executives belonged to diverse industries with annual revenues ranging between $ 500 million to $ 5 billion.

Based upon the organizational survey, the Saxon group came up with the following findings,

(i)   Only 25% of the information reported to the top management was predictive of future whereas 75% of the information was historical in nature pertaining to the internal operations of the organization.

(ii)   More than 80% of the enterprises surveyed did not employ an integrated performance management system. Further less than 10% of the enterprises which PMS used it for more than 5 years.

(iii)   Only 30% of the organizations had a detailed plan in place that clearly defined the expected future plans of the organization.

(iv)   For many of the surveyed enterprises tactical plans failed to describe the major initiatives to be taken up by the firm.

(v)   The knowledgable workers of the organization spent less than 20% of their time towards business intelligence and high value decision support tasks.

(vi)   The top management of the surveyed organizations did not provide enough time towards reviewing of the results from a strategic point of view.

(vii)   In case of most of the enterprises, there was no alignment between the organizations strategy, tactics and the expected outcome. Thus it proved that in most organizations gauging and analyzing of business performance was not provided enough weightage.

(viii) The financial records of many enterprises did not show a detailed cost benefit analysis report.

Based on the above findings by the Saxon group, it may be concluded that although business analytics and measuring of organizational performance are significant for the organizational survival, they are not being provided the due weightage.

3.4 PERFORMANCE MEASUREMENT

Q8.   What is performance measurement system? What are the characteristics of an effective performance measurement system?

Answer :   *Model Paper-I, Q8(b)*

Performance Measurement System

Performance Measurement System (PMS) is an important part of business performance management. According to professor Simons (2002), PMS can be defined as "A system which assist managers in tracking the implementations of business strategy by comparing actual results against strategic goals and objectives. A performance measurement system typically comprises systematic methods of setting business goals together with periodic feedback reports that indicates the progress against goals".

Thus, it may be understood that PMS helps the management with comparison of the actual target achieved against the strategic goals of the organization.

Characteristic Features of an Effective PMS

(a)   A good PMS would facilitate the management to align the top level strategic objectives with the grass, root level initiatives.

(b)   It would assist the enterprise to identify potential opportunities and threats before hand and help them to take proactive action accordingly.

(c)   It would guide the management to enhance the various organizational processes and procedures as per the need of the changing business environment.

(d)   It would assist the planners with forecasting and planning of the organizations strategic long term and short term goals in a timely manner.

(e)   It would provide a clear road map about the priority areas of the organization that needs to be financed on a priority basis.

(f)   It would facilitate the managers to delineate organizational responsibilities and prepare an objective reward and compensation plan which is in alignment with the roles and responsibilities of the different jobs.

(g)   A important characteristic feature of a good PMS is that it would make it easy for the managers to measure the actual performance of the employees with the expected performance. Thus, make them understand the extent of deviation and initiate corrective action.

Therefore, it may be stated that a holistic PMS is an integral part of any successful business organization. The different types of PMS being used by organizations since the last 40 years include. Activity Based Costing (ABC), Balanced Score Card (BSC) etc.

Q9.   Discuss KPI's in detail.

Answer :

The strategically aligned metrics of the organization are denoted as KPI (Key Performance Indicators). The KPI refers to a strategic objective of the organization and it is used to measure the performance against the strategic goals of the organization.

In the year (2009), researcher Eckerson stated that KPI's are multidimensional in nature and they exibit unique characteristic features such as,

(a)   Targets

The KPI's measure the actual performance against specific set targets. These targets may be of different types such as achievement targets, cost reduction targets, absolute targets etc. These organizational targets are defined in the strategic planning or the budget session.

(b)   Benchmarks

The targets are measured against certain benchmarks. This would help the enterprise to compare and analyze the actual performance against a predetermined benchmark. For example, the sales figure of the previous year may be considered as a measurement benchmark for comparing the current years sales performance.

(c)   Time Range

The performance targets are generally assigned a specific time frame/range by which time the target needs to be accomplished. Long term targets may be subdivided into smaller time targets which would act as a milestone tracker and help to keep the workforce motivated.

(d)   Strategy

Every KPI need to have a clearly defined strategic objective.

(e)   Ranges

The performance targets need to have a clearly defined target range.

Example

Between xxx and xxx range, below xxx range, above xxx range ect. Further these ranges can be encoded in the computer using different colours. Such visual display of the target range would make the understanding more user friendly.

KPI's may be sub-divided into,

(a)   Outcome KPI's

Also known as lagging indicators, these KPI's measure the performance of historical information. These are generally financial in nature.

(b)   Driver KPI's

These KPI's are also known as leading indicators or value indicators. These KPI's measure the key growth drivers of the organization. Example: Sales leads etc.

In some organizations, driver KPI's are called as operational KPI's. These KPI's include the following areas,

(i)   Sales Performance

Metrics of sales meetings, enquiry conversations etc.

(ii)   Service Performance

Service call resolution rate, service level agreements, performance delivery.

Q10. State the problems/drawbacks with existing performance measurement systems? What are the ingredients of a good collection of performance measures?

Answer :

Problems/Drawbacks with Existing PMS

Following are the major problems/drawbacks with the existing performance system,

(a) Most companies claim to have a good performance measurement system as opposed to performance management system.

(b) A research survey had found that 50% to 90% of the organizations use some variation of the Balance Score Card (BSC) for PMS. However when the managers were questioned as to what constitutes "Balance" in their BSC, they were unclear and confused.

(c) A major limitation is that more than 75% of the organizations focus upon performance measures which are financial in nature. Further their area of focus is primarily internal performance rather than external performance.

(d) Using only financial data for performance measurement suffers from the following limitations,

   (i)   It does not tell about future events.

   (ii) No clarity of the processes is provided by it.

   (iii) These focus upon short term achievement rather than long term goals.

(e) Many organizations state that they are tracking more than 200 performance measures at the organizational level.

   However, this would only complicate and confuse the end users rather than assisting in PMS.

(f) In case of many measures the management lacks direct control over the measure. For example, an organization can only monitor and analyze the extent of customer satisfaction, earnings per share, share price etc. The only areas of performance measure which can be fully controlled are employees actions. However, such control does not have a substantial impact upon the corporate level strategy or business level strategy. These would be effective only when the corporate goals are channeled down untill the bottom level of the organizational hierarchy.

Ingredients of a Good Collection of Performance Measure

A good performance measure would include the following ingredients,

(i) The performance measure need to focus upon key factors.

(ii) The measure need to be balanced in outlook, taking into consideration the past, present and future information.

(iii) The measures need to flow in a the top down hierarchy from the top till the bottom level of the organization.

(iv) The performance measure targets need to be formulated, based on scientific research and objectivity rather than arbitary.

(v) The performance measures need to balance the need of all the stakeholder's such as the shareholders employees, suppliers, and other key stakeholders.

3.5 BPM METHODOLOGIES

Q11. What are the two widely used approaches that support the basic processes underlining BPM?

Answer :   *Model Paper-II, Q8(b)*

Over the past 40 years, organizations had been using different systems for objectively measuring business performance. However the two most widely used approaches/systems of BPM are,

1. Balance Score Card (BSC)

The concept of balanced scorecard (BSC) was primarily developed by Kaplan and Norton in an article in the Harvard Business Review. The balanced score card - measures the drive performance'. It developed great levels of interest for the senior business managers which resulted in the further development. The attention was basically diverted from short-term measurement to creation of growth, learning and value-added services to the consumers.

Several organizations are making of the balanced scorecard as a supporting or underlying structure for making vital process decisions. The performance measurement process has been converted into strategic management system due to the recent development of balanced scorecard.

Balanced scorecard is defined as a conceptual framework which is used for converting the strategic objectives of an organization into a group/collection of performance indicators which are divided into four areas i.e., financial, customer, internal business processes and learning and growth.

Purpose of Balance Scorecard

Kaplan and Norton have found that balanced scorecard is used in many organizations in order to,

1.   Update and clarify the existing strategy.

2.   Recognize and arrange strategic initiatives.

3.   Communicate the strategy all over the firm.

4.   Coordinate both unit as well as individual goals with strategy.

5.   Connect strategic objectives to annual budgets and long term targets.

6.   Organize periodic performance reviews so as to learn and improve the strategy further.

2.   Six Sigma

Six sigma is a process improvement method which was developed by Motorola in 1980's. If a business is capable of producing only 3.4 defects out of every million opportunities, then the business is said to have met the "six sigma criteria".

In other words, six sigma is a statistical control limit given to the process of completing a job or task either completely or partially with a confidence level of 99.9997%.

Features of Six Sigma

1.   Six sigma aims at continuous improvement by reducing the cost of production and increasing customer satisfaction and return on investment.

2.   One of the features of six sigma is to produce the output defectless or error free.

3.   It ensures quality products.

4.   Six-sigma's philosophy is "Do it right the first time and every time".

5.   Six sigma's principles are applied in manufacturing and service industries.

6.   Six sigma is a statistical process control technique which is applied to gain complete confidence in the company's product and services and also the management.

7.   It is helpful to solve problems in an organised manner.

8.   The six sigma emphasizes on the following areas,

     (a)   Independent variable to the process.

     (b)   Eliminates the root cause of any problem and tries to prevent it.

     (c)   Input is given much importance than output.

     (d)   Emphasizes problem and not the cause.

     (e)   Concentrates on controlling the problem and not on monitoring.

Q12.   How does a BSC align strategies and actions?

Answer :

The Balance Score Card (BSC) facilitates the organization to align its daily actions with its long term strategic goals. Researchers Kaplan and Norton had laid down the following steps for implementation of BSC in the organization, which inturn would facilitate the alingnment of strategies and actions.

In order to implement a balanced scorecard approach within an organization, Kaplan and Norton have suggested eight steps.
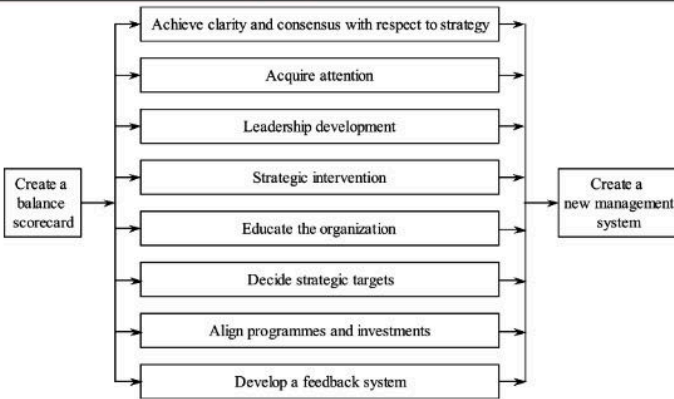
Figure: Kaplan and Norton's Steps for Implementation

The detailed approach of Kaplan and Norton can be successful by taking into consideration the three important aspects which results in the progress of a performance management system inclusive of the balanced scorecard. The three important aspects are explained as follows,

1.   Rigour in Purpose

On the basis of business objective, different industries hold different metrics. These metrics must be achieved in accordance with the company goals and four inter-related view points of the balanced scorecard. During a pilot exercise, metrics need to be clearly explained, validated and accepted by users.

2.   Rigour in Measurement

The success of the derived metrics would be based on the efficiency of gathering data and controlling the systems. This may differ from manual process on a spreadsheet to a difficult data warehouse.

3.   Rigour in Application

An organized and systematic, monitored balanced scorecard would be worthless if in case the data is not being utilized in enhancing and maintaining the performance. A process must be used in order to review the metrics constantly and take appropriate decisions for enhancing the performance. All the measures must have a target both for the present year as well as the 'best in class' for the future. In order to show the correct utilization of balanced scorecard, these measures need to be changed or altered.

Unlike other BPM systems, BSC utilizes two innovative tools for aligning strategies with actions,

(i)   Strategy map

(ii)   Strategic theme.

(i)   Strategy Map

A strategy maps describes in detail the steps of value creation for the organization utilizing a series of cause and effect relationships aligning them with major BSC elements like finances, customers, processes and learning and development. BSC and strategy map facilitate organizations to clearly describe, communicate and measure their strategies and actual actions.

(ii)   Strategic Theme

This tool divides the organizations strategic goal into sub themes or sub goals or value creation processes. Each of these strategic theme is a more simplified form of the strategic goals. Thus, by simplifying and breaking down these major goals into smaller parts, it makes these goals more manageable and increases the probability of success.

Q13. Write about DMAIC performance model and lean six sigma.

Answer :

DMAIC Method

1.   D - Define Phase

The define phase helps in defining the problem clearly as and when they occur. This is the first step in six sigma methodology. Define phase identifies those factors which are to be measured, analyzed, improved and controlled for greater revenue and needs improvement to carry out the project successfully.

Define phase should be stated clearly so that the problem can be solved effectively.

2.   M - Measure Phase

Measure phase is used to identify the critical internal processes which directly affects the CTQ measurements. Once the problem is defined, then the measure phase starts wherein it measures or evaluates the problems in the process which affects the CTQ standard.

The important factors of six sigma are, defect, variation, Critical-to-Quality (CTQ), process capability and Design for Six Sigma (DFSS).

After measuring the defects the impact of these defects should be measured because defects are measurable factors of a process. Once the defects are identified and measured the cost involved in them can be saved.

The defects in the process are analyzed by black-belt which affects the CTQ characteristic. The black-belt uses measurement systems analysis which consist of gauge studies and a complete evaluation of the capability of the process.

Gauge

The primary objective of gauge is repeatability and reproducibility study (gauge R&R). It is used to know whether measurement systems are functioning properly or not.

(a)   It helps in finding out the defects in the defined measurements and in the initial stages itself and helps in correcting them at the right time.

(b)   It provides information which forms the basis for determining the credibility and the areas to minimize defects.

Under the following four criteria gauge R&R repeats measurements,

(a)   Accuracy

To know how accurate the measurement is.

(b)   Repeatability

If the same item is measured many times will it give the same result?

(c)   Reproducibility

If the same items are measured by other people/other pieces of equipment, what will be the result?

(d)   Stability

To know the stability of accuracy, repeatability and reproducibility.

3.   A - Analyze Phase

Under this phase the defects which are affecting the Critical-to-Quality (CTQ) are analyzed. Hypothesis and statistical test are used to determine the factors which are critical to the outcome. Analyze phase acts as a cycle as it passes through various steps.

Step 1

Develop hypothesis about the reason(s).

Step 2

Analyze process and/or data.

Step 3

If the hypothesis is correct, add cause(s) to the list of vital few. If the hypothesis is incorrect refine it and again go to step 2 or reject it and go to step (1).

Hypothesis testing is used to compute the probability of those identified factors which have an impact on critical-to-quality outcomes. Soon after this, statistical conclusions are made for arriving at the corrective actions.

4.   I - Improve Phase

Once the factors which affect the variables are measured and analyzed then improve phase begins to rectify the defects. In improve phase, all those factors are improved which affect the critical-to-quality outcomes. Thus, the maximum acceptable range of each variable is identified and variations are measured using the measurement system.

5.   C - Control Phase

Control phase is used to maintain continuous improvement and in some cases control phase does not exist because mostly the problem gets eliminated in the above phases. Control phase ensures quality, productivity and improvement in the processes continuously. The scope of control mechanisms is wider.

Lean Six Sigma

In recent years organizations had been working towards integrating six sigma system with manufacturing this alingment is popularly known as lean manufacturing/lean production.

Lean Production

Lean production is also known as lean manufacturing. It is a systematic approach that identifies and eliminates the waste through continuous improvement by flowing the product as per the customer demand. Lean production aims at reducing the time, resources and efforts without making any compromise on the quality of the product.

The concept of lean corresponds to "no-waste", wherein waste refers to anything that is added to the cost of the product but not to the value of the product. There exist three types of works in any process. They are,

(a)   Value-Added

These are the process steps that produce value for the customer and for which the customer is ready to pay. In short, these are essential to meet the requirements of customer.

(b)   Non-Value Added

These process steps are not essentially needed to produce value for the customer but are required for other reasons such as fulfilling company mandates, legal requirements and regulatory requirements.

(c)   Waste

The process steps that produce no value for the customer but consumes resources.

The company 'Toyota' originated the concept of lean production. It identified eight types of waste,

1.   Defects

Time and money wasted for determining mistakes and fixing the defects.

2.   Over-production

Production more than what is needed.

3.   Waiting

Work delayed due to waiting machines, people and material.

4.   Transportation

Time and efforts involved in movement of materials, products, information and people.

5.   Motion

Unnecessary work-hand offs or transfer.

6.   Extra-processing

Doing more work on improving the quality than needed.

7.   Inventory

Holding additional material or the resources that are not required.

Q14. How six sigma success can be increased? Discuss.

Answer :

Six sigma is like any other business initiative. It need to nutured, enabled and maintained on a regular basis to increase its success impact. According to researcher Richardson 2007, "Six sigma is not the end all be all. It is simply a set of process tools we could never suggest that the company's performance is solely linked to the adoption of these tools".

The following are the six sigma enablers,

1.   Leadership Commitment

For making the six sigma successful the firm must receive a total commitment from the top management and senior leadership. In GE, Jack Welch and Bob Galvin gave their total commitment for making the six sigma successful.

2.   Change in Culture

Six sigma mainly focuses on the predetermined tangible and time bound objectives of the business, the culture that is changed due to the changes in the perspective of delivering the tangible or positive results.

3.   Change in Attitudes

Six sigma ensures a complete change in the attitudes of the management and the organizational employees. All the members have to change themselves and should focus on the attainment of the final objective by using the continuous improvement process which helps in detecting and eliminating the defects.

4.   Challenging Standards

The six sigma develops challenging standards which are complicated but are very effective. The attainment of these standards helps the firm to attain improved organizational performance.

5.   Customer Flows

Six sigma is a statistical measure of process which helps in attaining the requirements of the customers. The main objective of the of six sigma process is to fulfill the needs of the customers effectively. It examines the processes and process measures from the customer's perspective.

6.   Continuous Improvement

Six sigma consolidates the processes and regular work and implements them in such a way that continuous improvements are made for achieving the final objective of satisfying the customers.

Q15. Distinguish between balanced score card and six sigma.

Answer :

Following are the difference between balance score card and six sigma,

	Six Sigma		Balance Score Card
1.	Six sigma possesses higher quality of products or services which holds the probability of only 3.4 defects per million opportunities (DPMO).	1.	Balance score card relates to the strategic goals and objectives of the business organization.
2.	Six sigma is a problem solving methodology which helps in reducing defects and its related costs.	2.	Balance scorecard has been designed and developed to balance the organizations strategic goals. It focuses towards vision and values.
3.	Six sigma considers the process management, improvement, measurement as daily activities for improving the process management.	3.	Balance score card considers the strategic management system measures.
4.	Six sigma not only produces defect-free goods or services or continuous improvement but also helps in reinvention/ reinnovation of the business.	4.	Balance score card not only identifies measurements around the organizations vision and value, but also highlights targets for each measure.
5.	Six sigma's implementation in the company brings standard improvement in the performance.	5.	Balance score card implementation helps with identifying measures for key management processes such as organizational vision, strategy, plan, feedback etc.

6.	Six sigma uses 'critical-to-quality' aspect which is most crucial from the customer's point of view.	6.	Balance score card uses elements such as organization's internal operations, strategic goals, customers and so on.
7.	Six sigma also solves the customer's problems effectively.	7.	Balance scorecard helps towards balancing strategic organizational elements, such as customers, internal operations etc.
8.	In six sigma, Champions, black belts, master black belts and green belts are used for successful implementation of six sigma projects.	8.	Balance score card employs strategy maps, strategic themes and other tools for successful implementation of its mission in the organization.
9.	The ISO, CMM and six sigma are integrated together in the system in order to improve the efficiency of the services which helps in solving the problems.	9.	Balance score card is an integration of performance management elements and management methodologies to facilitate the enterprise to achieve tartets related to its finances internal processes, customers and learning and development.
10.	Six sigma follows, "do it right the first time" philosophy.	10.	Balance score card follows the goal of translating vision and strategy into clear objectives, measures, targets and initiatives.

3.6 BPM ARCHITECTURE AND APPLICATIONS

Q16. What is logical system architecture? What are the basic parts/elements of a BPM architecture?

Answer : *Model Paper-III, Q8(a)*

Business performance management is a wider term which covers many processes, metrics, tools, techniques and methodologies used by the organization to measure, monitor and manage its performances on a regular basis.

The concept of logical system architecture refers to the logical and physical design of the system. Logical system includes the functional aspects and their interactions whereas the physical system explains about how the logical design is implemented.

Elements/Basic Parts of BPM Architecture

The BPM system consists of the following three layers,
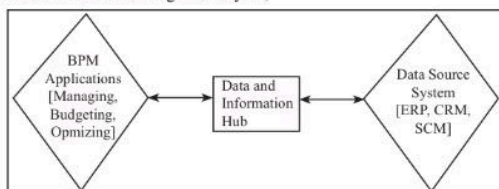
Figure: Showing BPM Logical System Architecture

(a) BPM Applications

This layer of BPM architecture entails BPM process which transform user interactions and source data into plans, budgets, reports etc. There does not exists a "One size fits all" BPM application. Depending upon the specific strategic needs of the organization the BPM application would differ from one organization to the other.

(b) Data and Information Hub

The BPM system needs information inputs, from multiple sources such as spread sheets, external data CRM, ERP etc. However in well structured and designed organizations, the variety of different information from multiple sources are stored in a central data base location such as data ware house.

(c) Data Source System

This layer of BPM architecture entails all different data sources for the enterprise. These includes sources such as ERP, SCM, CRM, spread sheets external info logs etc. In case of large business houses, financial and operational data is assembled from multiple organizational systems.

Q17. Discuss the major categories of BPM applications.

Answer :

According to the analyst group Gartner incorporation (2009) majority of the closed loop processes can be handled by the following applications,

(i) Strategic Management Application

These applications help support elements such as monitoring of corporate performance, strategic planning, strategic modeling, quick decision making etc. These applications are aligned with strategic maps and methodologies such as balanced scorecard.

Further strategy management application can provide capabilities for,

(a) Scorecards and strategy maps for gathering the strategic goals, measurement of the major elements of organizational performance, enterprise wise communication and so on.

(b) Dash boards to gather and display metrics.

(c) Goal management tools.

(ii) BPF

These applications assist in budgeting, planning and forecasting (BPF). These focus upon short term financial goals, long term and strategic goals as well. Further they need to encourage the creation of an enterprise wise planning model.

(iii) Optimization Modeling Applications

These applications focus towards optimizing the resource allocation and boosting profitability of the organization. Activity Based Costing (ABC) is a good example of this, application. Although the main focus area of these application is costing of resources, some new applications had moved beyond the traditional area and are also dealing with allocation of costs for channel strategies, packaging, bundling and so on.

(iv) Statutory and Financial Applications

BPM applications are also needed to align themselves with international accounting standards such as GAAP (Generally Accepted Accounting Principles) etc. These also comprise of visualization techniques which are specifically designed to support variance analysis from the budgets or targets.

(v) Financial Summarizing Applications

These applications helps enterprises to consolidate and summarize a variety of financial data taking into account the accounting standards in place and also the legal provisions prevailing in the country. These application form an integral part of the Business Performance Measurement system as they facilitate in the creation and audition of financial data and information which would be used by other BPM applications to compare variation (if any) from the planned target.

Q18. What changes are observed in the BPM market in the last 3 to 4 years? What are the categories in gartners magic quadrant? Who are some of the vendors in the BPM leaders quadrant?

Answer :

The BPM market is highly dynamic and competitive in nature. It comprises of software enterprises which provide core BPM applications such as budgeting, planning, forecasting, score carding etc. There had been many changes observed in the BPM market during the past 3 to 4 years as discussed below,

1. The biggest change was the consolidation of the BPM market vendors. For example, previously there were multiple vendors such as hyperion, SAS, cognogs, etc. However, these smaller to BPM vendor had been acquired by bigger software firms, such as,

(a) Cognos vendor was acquired by IBM.

(b) Hyperion vendor was acquired by oracle incorporation.

2. According to gartner, the value of commercial BPM market had increased to 2 billion dollar by the year 2007 and the estimates predicted a very high increase over the next few years.

3. The main driver for this growth is that most firms are replacing spreadsheet based applications with more advanced BPM softwares.

4. There had been a very big consolidation in the BPM market, for instance. IBM, orace, SAP and SAS together account for 78% of the total global BPM market.

5. Gartner had formulated a magic quadrant for positioning BPM companies based upon various parameters.

Categories in Gartners Magic Quadrant

Gartner had categorized BPM vendors based upon their ability to successfully execute tasks and the strength of the vendor organizations vision. Figure Garnters magic quadrant.
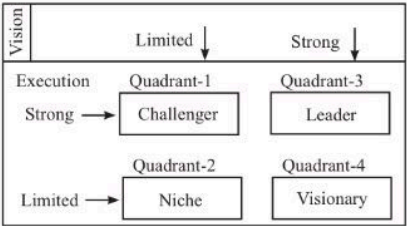
Thus, based upon the above shown gartners magic quadrant, vendor firms would be categorized as challenger, niche, leader or visionary.

Some of the BPM vendors which fit under the leader category are oracle incorporations, hyperion, SAP business objects and IBM's cognos. The reason for these vendors to be categorized as leaders is the sheer size of these major vendors.

3.7 PERFORMANCE DASHBOARDS

Q19. What do you mean by dashboards? Differentiate between dashboards and scorecards. What layers of information are provided by the dashboards?

Answer : *Model Paper-III, Q8(b)*

Business performance management systems involves systems like, dashboards and scorecards are the most commonly used BPM systems by organizations.

Dashboard

The dashboard system displays vital information to the top management in the form of graphs, pie charts, barcharts or other visually appealing instruments on a single screen. This possess all vital business information in a user friendly manner which will help the top management to compare, contrast and analyze vital information and then take proper decisions. An illustration of a typical dashboards is shown below,

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.
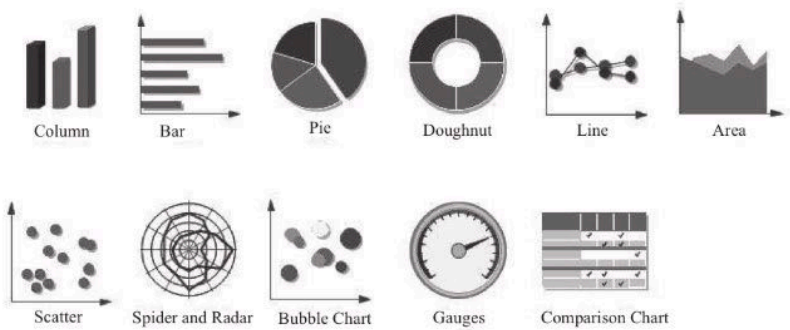
Difference between Dashboard and Scorecard

In most BPM journals and magazines, dashboards and scorecards are used interchangeably because both consist of similar features. However, the main point of difference between the two is that, dashboards are used by managers for making operational and tactical decisions, sometimes on a weekly, daily or hourly basis, whereas scorecards are employed to monitor the strategic alignment and to measure the extent of success achieved with regard to the strategic objectives and the targets.

Information Provided by Dashboards

According to BI specialist Dr Eckerson (2016), a typical dashboard provides these layers of information,

(a) Monitoring Data

A dashboard provides information in graphical or abstract form for monitoring the organization's key performance metrics.

(b) Data Analysis

It summarises and provided information in easily understandable or user friendly format, which would facilitate in the identification of the root cause of any problem.

(c) Decision Data

An important layer of information provided by the dashboard includes detailed operational information in graphical dimensional form, which would facilitate in the easy and quick decision making process. The best feature of a dashboard is that it provides all the vital information on a single screen. According to BPM analyst, Mr. Few (2005). "The fundamental challenge of a dashboard design is to display all the required information on a single screen clearly and without distraction in a manner that can be assimilated quickly".

Q20. What are the common characteristics of dashboard and scorecard?

Answer :

A dashboard provides vital organizational information to the managers on a single screen. A scorecard and dashboard are similar in outlook but not the same. According to researcher Novel (2009), the common characteristic or features of dashboard and scorecard are as follows,

(i) Both these BPM tools are visual in nature. They summarize and present the required information on a single computer screen.

(ii) Both these BPM system employ a variety of user friendly, compact and concise elements such as charts, graphical information, flow charts, performance bars, spotlights, sparklines, different kinds of gauges, pie charts, pop-up visual tools and so on.

(iii) Both dashboard and scorecard are highly user friendly and transparent in nature. This means that the end user does not require any specialized training to read and understand their contents.

(iv) They possess the characteristic feature of unifying information from multiple BPM system into a single platform.

(v) These BPM system are equipped with the feature to perform a high percision drill down program and drill through program with the help of very large data base of information or reports. It provide comparative and evaluative reports to the managers at a very high speed.

(vi) These BPM systems provide refreshed and uptodate information on a real time basis. This enables the end users to make informal decisions with most updated information on hand.

(vii) No separate coding or customization is needed to deploy and maintain dashboards or scorecards.

(viii) They employ colours and high lighters (red, blue, green etc.) in an appropriate manner. This would help to attract the managers attention on the most vital part of the summarized information that is presented on these BPM systems.

(ix) Dashboards and scorecards make use of specialist display widgets or other unique features to highlight comparative and evaluative reports to the end users.

(x) Performance dashboards and the commonly used standard performance scorecard possess the ability to align well with the larger performance measurement systems which would be used by any business organization.

SHORT QUESTIONS AND ANSWERS

Q1.   Write briefly about TSA's PIMS?

Answer :   *Model Paper-I, Q3*

The U.S based transport organization TSA (Transport Security Administration) employs a Business intelligence software called PIMS (Performance Information Management System). This system is a perfect example of the ability of BI systems to be strategically oriented and centrally controlled.

The PIMS performs multiple tasks such as tracking of passenger lines, keep an updated record about overtime, absenteeism injury to passengers (if any) etc. This highly efficient and successful. PIMS is designed on microstrategy.com, BI software and its operations are critical for the smooth running of the daily services. Using the PIMS system had saved the transport organization more than 100 million dollars.

Q2.   Why a company needs a well formulated strategy?

Answer :

The following points explain the need for strategic management,

1.   Give Guidelines

Strategic management give guidelines to employer regarding what organisation expecting them to do. This guideline make employers clear about what their jobs are demanding and what performance they are supposed to give. Strategic management offer incentive to employer and assist the organisation in the attainment of objectives.

2.   Chances of Good Performance

Most of the research studies states that there exists a relationship between good performance and formal. Planning and companies which do strategic planning will have more chances to succeed than companies who don't do strategic planning.

3.   Facilitates Communication

Strategic management facilitates free flow of communication of information form bottom level managers to middle level managers to top level managers.

4.   Allocation of Resources

Strategic management assist in selecting achievable and realistic projects and inturn enhances the allocation of resources to realistic projects.

5.   Allows Forecasting of Change

Change is stable or constant and it makes planning a complicated task. Companies might pro-act to the changes in the environment instead of responding to change. Strategic management motivate top level managers to forecast the changes in the environment to direct and control them. Apart from this, strategic management permit company to grab the opportunities coming out of the changes in the environment and overcome threats by forecasting future change. Therefore, strategic management permit firms to take decisions based on long-term forecasts.

Q3.   State the key elements of a diagnostic control system?

Answer :   *Model Paper-II, Q3*

A diagnostic control system is a cybernetic system, that is, it intakes inputs, process the information and provide outputs. This output is then compared with pre-established standard performance benchmarks and feedback is provided stating whether there had been any variance from the expected performance.

All organization, irrespective whether they employ BPM system or not, would use some type of diagnostic control system. The key elements of this system are,

(a)   Inputs

In the form of actual performance.

(b)   Monitoring Elements

This include any type of tool, technique, metric or measure used by the organization for processing the information. Example: Balanced Score Card, Dashboards, Human Resource System, Financial management system etc.

(c)   Output

This element comprises of any type of end result produced by the monitoring system.

Q4.   How does BSC and six sigma can be integrated?

Answer :

The BSC (Balance Score Card) and six sigma system can be integrated by making the following changes in the organizations.

(a)   The organization strategy need to be converted into quantifiable goals. This aim can be achieved by the employment of scorecard monitor and mapping of the organizational strategy.

(b)   Enterprise level strategy need to be broken down into smaller parts and the aims mission and actions are passed down to the lowest level of the organization.

(c)   Integration of BSC and six sigma requires the formulation of key organizational goals based upon constructive feedback received from the customers.

Q5.   Distinguish between lean production and six sigma.

Answer :   *Model Paper-III, Q3*

Differences between lean production and six sigma are as follows,

	Area	Lean Production	Six Sigma
1.	Aim	The main aim/purpose of lean production is removal of wastage from the process of production.	The main purpose of six sigma is reduction of variance and elimination of errors.
2.	Focus	The key focus area is the production flow.	The key focus area is the cause of variance (problem).
3.	Approach	Its approach is to work towards many small improvements.	Its approach is to remove the root cause of the problem.
4.	End Result	Its end result is the successful elimination of wastage and increase in the production efficiency.	Its end result is less variance in output and substantial reduction in the errors during performance of services.

Q6.   Write a note on data visualization.

Answer :

Professor Few (2008) had defined data visualization as "the use of visual representation to explore, make sense of, and communicate" data (to the user). It is related to the field of information technology, information graphics, scientific visualization and stastical graphics.

In simple terms, data visualization means presentation of any data in visually appealing and easy to understand format. Balanced scorecards and dashboards are the best examples of systems that deal with visualization of data.

It can be presented in the form of charts, graphs, spotlights, gauges, highlighters, banners etc.

INTERNAL ASSESSMENT

I. Multiple Choice

1. Lean production is also known as _____.   [   ]

    (a) Job production

    (b) Batch production

    (c) Lean manufacturing

    (d) Slim manufacturing

2. Different tools, techniques, methods and metrics used for measuring organizational performance are considered as _____.   [   ]

    (a) Business Performance Measurement (BPM)

    (b) Business Program Management (BPM)

    (c) Performance Management (PM)

    (d) Performance Management System (PMS)

3. BPM is _____ in nature.   [   ]

    (a) Static

    (b) Narrow

    (c) Both (a) and (b)

    (d) Dynamic.

4. Harrah's closed tool marketing model was presented by _____.   [   ]

    (a) Watson and Volonio

    (b) Watson

    (c) Volonio

    (d) None of the above

5. An effective performance measure need to focus upon _____.   [   ]

    (a) KPI's and Targets

    (b) KRA

    (c) Objectives

    (d) Standards

6. The philosophy "do it right the first time" is followed by _____.   [   ]

    (a) Balance score card

    (b) Dashboard

    (c) Six sigma

    (d) None of the above.

7.   The use of visual representation to explore, and communicate data is _____.   [   ]

      (a) Infographics

      (b) Data visualization

      (c) Data graphics

      (d) Data visuals

8.   Gartner's magic quadrant consist of all the elements except _____.   [   ]

      (a) Unlimited

      (b) Strong

      (c) Vision

      (d) Limited

9.   The _____ system displays vital information on a single screen.   [   ]

      (a) Dashboard

      (b) BSC

      (c) Both (a) and (b)

      (d) None of the above

10.   A common feature of BSC and dashboard is,   [   ]

      (a) Single screen

      (b) User friendly

      (c) Data visuals

      (d) All the above.

II.   Fill in the Blanks

1.   _____ is an important part of business performance management.

2.   An example of data source system is _____.

3.   DMAIC stands for _____.

4.   BSC stands for _____.

5.   KPI stands for _____.

6.   _____ refers to the difference between the organization strategic plan and the actual execution of the plan.

7.   SWOT stands for _____.

8.   BI stands for _____.

9.   ABC stands for _____.

10.   DPMO is the abbreviation for _____.

KEY

I. Multiple Choice

1. (c)

2. (a)

3. (c)

4. (a)

5. (a)

6. (c)

7. (b)

8. (a)

9. (c)

10. (d)

II. Fill in the Blanks

1. Performance Management System

2. Enterprise Resource Planning (ERP)

3. Define Measure Analyse Improve Control.

4. Balance Scorecard

5. Key Performance Indicators

6. Strategy Gap

7. Strengths Weakness, Opportunities, Threats

8. Business Intelligence

9. Activity Based Costing

10. Defects Per Million Opportunities.

III. Very Short Questions and Answers

Q1. Define BPM?

Answer :

Business Process Management can be defined as a sum total of tools techniques and matrices used to measure and manage the organizational performance. It is also called as corporate performance management, or enterprise performance management.

Q2. Write about Harrah's closed-loop marketing model.

Answer :

Harrah's closed loop marketing model was presented by Watson and Volonio in the journal "Harnessing customer information for strategic advantages technical challenges and business solution" (2001). The main aim of this model is to analyze the customer responses and adjust the business offering accordingly, to run the business successfully.

Q3. Write any two features of an effective PMS.

Answer :

(a)   A good PMS wood facilitate the management to align the top level strategic objectives with the grass, root level initiatives.

(b)   It would assist the enterprise to identify potential opportunities and threats before hand and help them to take proactive action accordingly.

Q4. Write the purpose of BSC?

Answer :

Kaplan and Norton have found that balanced scorecard is used in many organizations for following purposes,

1.   To update and clarify the existing strategy.

2.   To recognize and arrange strategic initiatives.

3.   To communicate the strategy all over the firm.

Q5. What do you mean by six sigma.

Answer :

Six sigma is a process improvement method which was developed by Motorola in 1980's. If a business is capable of producing only 3.4 defects out of every million opportunities, then the business is said to have met the "six sigma criteria".

UNIT 4

Business Analytics and Data Visualization

LEARNING OBJECTIVES

After studying this unit, one would be able to understand,

* ❖ The Definition of Business Analytics (BA).
* ❖ Tools of Business Analytics.
* ❖ Techniques of Business Analytics.
* ❖ Advanced Business Analytics Tools.
* ❖ Concept of Web Analytics.
* ❖ Uses of Business Analytics.
* ❖ Success of Business Analytics.
* ❖ The concept of data Visualization and new directions in it.
* ❖ Purpose and usage of Visualization Spreadsheets.
* ❖ Concepts of GIS and GPS.

INTRODUCTION

The subject of Business Analytics revolves around the concept of decision making. Every individual makes decisions in life. These may be related to purchasing a specific item or performing or not performing a specific actions. In the business world, enterprises need to make strategic decisions, which would have a long term impact upon the organization.

Vendors also provide exclusive web support options to the business enterprises for them to integrate there BA technologies with the web. For example Google provides "Google Analytics" software for medium and small enterprises.

Web analytics refers to the integration and applicability of the business analysis activities with web based processes.

Data visualization is the concept that refers to the technologies supporting visualization and data interpretation at various instances in data processing. It involves digital images, GIS, Graphical user interfaces, Virtual reality, Graphs, Videos, Animation and dimensional presentations. It can be further more applied to mainstream computing and intelligent visualization. Visual spreadsheets are the tools used by the end users for programming decision support applications.

Graphical Information System (GIS) is a computer based system which enables us to capture, store, analyze and display the data geographically from a base location. Global positioning system is a space based weather radio navigation system by which the time, location and velocity of an object can be determined any where on the globe at any time.

4.1 BUSINESS ANALYTICS – DEFINITIONS AND PREREQUISITES

Q1. Define analytics and Business Analytics (BA). State the prerequisites for effective business analytics.

Answer :   *Model Paper-I, Q9(a)*

The present day business world is highly competitive and saturated. Globalization had increased the need for organization to be smart, innovative and accurate in predicting the customer taste, demand and so on. Therefore, organizations need to objectively and accurately analyse the business environment.

Definitions

Analytics can be defined as the art and science of analysing any data or information. It involves a detailed examination of the available content which is investigated and inspected with a specific motive.

Business Analytics (BA)

Business analytics refers to the broad category of skills, tools and techniques employed for gathering, storing and analysing of business related data and information by any business enterprise. The main aim of business analytics is to help the management to arrive at an objective and accurate business decisions.

Business analytics is known by multiple other names such as Business Intelligence, BI application, B.A processing and so on. BA has become a common tool which is used by many big and medium organizations to help them make prudent strategic decisions.

For example, popular fast food Chain "Pizza Hut" employs B.A tools to gather useful customer related information such as,

(a)   Which pizza, the customer would order mostly?

(b)   What time the demand would be high/low?

(c)   What kind of promotional campaigns would be most impactful?

Many banks use BA softwares which help them decide whether to provide a loan to the customer or not (depending on his/her credit worthiness which is computed using BA software).

The scope of BA is very big and encompasses areas of business financial budgeting, competitive intelligence, demand forecasting, customer data base management, supply chain management and so on.

Prerequisites for Effective Business Analysis

The prerequisites for effective business analysis are as follows,

(a)   Goal Statement

A goal/mission statement is the most important prerequisite for BA to be effective. Every organization (big or small) need to have a clear and precise mission statement, which would motivate and guide its members to strive and work towards the achievement of the specific mission statement. For example, the mission statement of I.B.M states "we strive to lead in the invention, development and manufacture of industry's, most advanced information technologies, including computer systems, software storage systems and micro electronics".

(b)   Top Management Support

It is necessary for the full support of the top management for the effective and result oriented implementation of BA.

(c)   Team Work

Business analytics team need to work in sync with other departments of the organizations for collecting the required data and information which would be used for preparing BA reports.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD. ──────────────

4.1.1 Tools and Techniques of Business Analytics (BA)

Q2. Discuss in detail tools and techniques of Business Analytics (BA).

Answer :   *Model Paper-II, Q9(a)*

Business analysis utilizes a large number of tools and techniques. These can be sub divided into,

(a) Information and knowledge tools

(b) Visualization tools

(c) Vendor classification tools

(a) Information and Knowledge Tools

These include the following tools,

(i) OLAP

Online Analytical Processing (OLAP) refers to different variety of activities which are performed by the end user over the Internet. These includes generation of queries, answering of queries, requests for ad hoc reports and graphs, building of visual presentations and so on.

(ii) Reports

Codd's rules state that reporting tools need to be uniform and flexible. The two most popular form of report tools are,

(a) Routine Reports

These BA tools are automatically generated and distributed on a periodic basis. For example, the business analytics software would generate a weekly sales report of the store which is auto generated and mailed to the concerned managers.

(b) Ad-Hoc Reports

These BA reports are created on demand for a specific user. For example, a report request for providing the list of customers who had purchased $ 10,000 or more worth of electronic item during the last three years.

(iii) Data Mining

It is one of the most widely used B.A tool. The data mining tool mines the database and extract hidden, predictive information. This would be helpful in key decision making.

(b) Visualization Tools

These BA tools are employed to have a balanced view of the organization. These includes the following,

(i) Balanced Scorecard

The concept of Balanced Score Card (BSC) was primarily developed by Kaplan and Norton in an article in the Harvard Business Review. The balanced score card - measures the drive performance'. It developed great levels of interest for the senior business managers which resulted in the further development. The attention was basically diverted from short-term measurement to creation of growth, learning and value-added services to the consumers.

Several organizations are making of the balanced scorecard as a supporting or underlying structure for making vital process decisions. The performance measurement process has been converted into strategic management system due to the recent development of balanced scorecard.

Balanced score card is defined as a conceptual framework which is used for converting the strategic objectives of an organization into a group/collection of performance indicators which are divided into four areas i.e., financial, customer, internal business processes and learning and growth.

(ii) Dashboards

These are visual display tools which displays raw data and visual graphics. This BA tool helps in measuring the performance.

(c) Vendor Classification Tools

For answer refer Unit-IV, Page No. 4.4, Q.No.3.

Q3.   Explain how vendors classify the tools of Business Analytics (BA).

Answer :

There are a huge range of business anatytics tools, techniques and software which can be employed by any business enterprise. Vendors classify BA tools into the following categories/styles.

1.   Enterprise Reporting Tools

These BA tools are highly effective for generating accurately formatted static reports. These can be used by multiple departments of the organization. These reports are pixel. Perfect in nature and are used for operational reporting.

2.   DAT

DAT stands for Delivery and Alert Tools. These are highly powerful distribution engines which are employed by organizations to send full sized reports or remind alerts to a large group of customers. These DAT's are based upon reports prepared from the customer database.

3.   Statistical Tools

These BA tools are used for analying the product demand, based upon multiple factors. These tools comprises of a variety of highly sophisticated mathematical and data mining tools which are used for carrying out predictive analysis/forecasting.

4.   CAT

Cube Analysis Tools (CAT) are the business analytics tools used by vendors to provide OLAP and simple slice and dice analytical information to business enterprises.

5.   Ad hoc Query Tools

These BA tools are used to get information by slice and dice of the entire data base. Under 'slice and dice' concept, the data is rearranged and view from different perspectives.

Many vendors also categorise BA tools into the following two categories,

(i)   EIS

Executive Information System (EIS) is a very common business analytics tools used by manager and business executives to analyze the organisational reports and related database to generate reports and statistical summary about the organisational performances. EIS softwares are highly user friendly and support a wide range of graphics and other related services.

(ii)   ESS

ESS stands for Executive Support System. The scope of ESS is much wider and indept, as compared to EIS. ESS provides the top management and executives the option of information intelligence support, office automation and other required capabilities.

4.1.2 Advanced Business Analytics

Q4.   Why are advanced analytics tools necessary for businesses? Explain data mining and predictive analysis as special tools for advanced business analytics.

Answer :

Need for Advanced Analytics Tools

The need for advanced analytical tool is required as a result of total integration of business and commerce with information technology. Every activity of any business organization is partially or wholly dependent upon the employment of information technology. Thus, it has become mandatory for organization to employ highly sophisticated business analytic tools to survive, remain competitive and increase their market share in today's highly competitive business world. Data mining tools and predictive analysis are two of the advanced business analytics tools used by global organizations.

Data Mining

The data mining tools efficiently retrieve hidden predictive information from the provided database. Further, this BA tool helps in searching for patterns in a large database. They employ highly powerful mathematical and statistical elements for mining of useful information from the database.

Data mining tools look similar to OLAP. However an important point of distraction between data mining and OLAP is that unlike data mining tools, OLAP can only provide answers to specific questions asked. It does not come up with significant questions and answers all by itself.

Predictive Analysis

These are some of the most advanced B.A tools employed by organizations. They help in predicting an outcome for a given event, many hours or many days in advance. This help the firms to take required actions accordingly.

For example, US based BA firm Inrix.com provides predictive solutions to its client organizations for a nominal fee. It is used by the road department for monitoring real time traffic flow and also the probability of traffic jam in the near future. Infact these advanced BA software tools are so powerful that they can predict by more then 90% accuracy probability of traffic jams for many days in advance. They use predictive algorithm and combine it with date to create predictive snapshot of expected events and provide highly useful information to the end users.'

Q5. What are the tools/products offered by vendors for advanced business analytics?

Answer :

The vendors provide a wide variety of highly sophisticated and accurate business analtytic products to organizations to facilitate correct and object strategic decision making. They include,

1. Microsoft Incorporation

This vendor provides a powerful BA software named, "Microsoft Dynamics GP 9.0". Which helps organizations to conduct advanced data analysis.

2. Micro Strategy

This vendor provides more than 400 different statistical, mathematical and financial tools for organizations to quickly create reports and also analyse data.

3. Cognos 8

This vendors software, "Cornos 8. Business intelligence" includes customized time series analysis tools, demand forecasting tools, drill down and data optimization options which could help the firms to become more competitive.

4. ILOG

This vendor provider good optimization suite, which would help the firm to maximize its resource utilization capabilities and create quick cost benefit analysis reports.

5. Fair Isaac

This vendor provides BA tools for accurately conducting fraud deletion, risk analysis, profitability analysis and so on.

6. SPSS

This vendor is a market leader in providing tools for predictive analysis and advanced data mining.

7. Other prominent vendors who provides good BA tools includes oracle, SAS, Statsoft, Knowledge Extractum Engine Angoss software and I.B.M.

4.1.3 Business Analytics and Web

Q6. Define Web Analytics. Describe the use of web in business analytics.

Answer :

Internet has revolutionized the way business is carried out globally. In the present day business world every aspect of the business is fully integrated with the web/internet. The different tools, techniques skills and softwares of business analytics also need to be integrated with the web to provide real time information to the executives.

Definition of Web Analytics

It refers to the integration and applicability of business analytics activities with web based processes. In simple terms web analysis refers to the interface between BA and the internet. It also includes the integration of business analytics along with e-commerce tools and techniques.

According to BA researcher Schlegel (2003), the methodology and tools of web analytics are highly visual in nature. He proposed the term click stream analysis.

Uses of Web in Business Analysis

A large proportion of business analytics tools and software technologies are related to the web. For example, Global heavy electronics manufacturer Hitachi Incorporation (Japan) employs a wide range of web based e-commerce tools which are fully integrated with business analytics. Software applications such as ERP system, data warehouses etc. Integration of B.A with the web helps to create a synergy between both the elements. It enriches the business analytics data by providing real time inputs gains through the various e-commerce and web technology softwares.

Many multinational organizations are coming up with special tools and softwares which are being integrated with web technologies to facilitate prudent decision making by the enterprise managers. For example, IBM uses a BA software known as decision edge software, which offers the end users OLAP capabilities on the Internet through the use of multiple web based search engines, browsers and other web technologies.

Another case in point is the BA software tool of Oracle incorporation. Its BA tools "Financial Analysers and Sales Analyses and Hummingbird BI can be early integrated with the latest web based technologies.

The scope of web analytics is huge and still contains untapped potential to be explored for the welfare of the organization.

Q7. Define and discuss click stream analysis.

Answer :

Click stream analysis refers to the analysis or study of the mouse clicks of the internet users computer. Thus, it refers to the analysis of the various web pages that an internet user had visited, the amount of time he had spent on that particular web page, the various other web pages that the customer had visited and for how much duration. Further it also includes the recording, storing and detailed analysis of the sequence in which the web user had visited the various web pages.

Thus, click stream analysis refers to the analysis of a variety of different data that is gathered from the internet/web environment. The data gathered for click stream analysis is known as click stream data.

The major benefits/uses of click stream analysis are,

(a) It provides the business enterprise an idea about the way the web user visit of different web pages.

(b) The interest of the potential customers can be forecasted by analysing the sequence in which the web user clicked upon different web pages. Further, the time duration spent upon each web page also indicates the interest of the end user (potential customer).

(c) It would help the enterprise to measure the effectiveness of its web based promotional ads.

(d) Based on the click stream analysis results, the firm may either continue with the existing web promotional campaign or it may create a new, more attractive web advertisement.

The data and information that can be gathered using the click stream analysis is huge. For instance, web analytic researchers Werner and Abramson described a method of sorting and aggregation using which the software engineers can process 1 billion web records per day and collect useful business analytical information.

However, successful performance of click stream analysis require the following,

(a) The raw click stream data need to be sourced from multiple sources.

(b) The analysers need to employ cookies and session identification numbers in the URL's for the identification of amount of the data session.

(c) A big proportion of the projects resources need to be allocated for date preparation.

Q8.   Discuss vendor support for web analytics.

Answer :

Many vendors such as Google incorporation, Cognos, web trends, informatics etc., provide extensive web support options to business enterprises to integrate their business analytics technologies with web technologies. These are discussed as follows,

1.   Google

This popular vendor provides a software. "Google Analytics" to small and medium enterprises for free of cost. This web technology can be used by business organizations for click stream analysis of web logs.

2.   Business Objects

This vendor provides its client organizations options to query and analyze web content through its software "Web Intelligence". It facilitates the client enterprises to track manage and store information in multiple web based data sources which can be later integrated as per the organizations conveniences.

3.   Cognos

This vendor provides its client organization its web software named "Cognos 8" which helps its client customers to develop softwares easily.

4.   Informatica

This vendor provides its client organization "The informatica Analytics Delivery Platform" which helps business enterprises to closely track their business performance. It provides the "Power Centre 8 BI Platform" though which organization can gather detailed business performance metrics.

5.   Web Trends

This vendor provides business organizations to analyse the internet traffic on a real time basis. Employing its, software products, organizations can track customer purchasing trends, revenues, effectiveness of sales, promol and other interesting and useful details, about the online customers.

6.   Advisor Solutions

It offers the client organizations an interactive chart library that provides useful information about business display needs, on demand real time analysis, point and click self services etc.

7.   Angoss Knowledge Studio

This vendor provides good web mining tools for client organization.

4.1.4 Usage, Benefits and Success of Business Analytics

Q9.   What is the usage of Business Analytics? What are the benefits of implementing Business Analytics in an Organization.

Answer :   *Model Paper-III, Q9(a)*

Usage of Business Analytics

In the present business scenario, the usage of Business Analytics increasing day by day. Most of the medium and large scale organization successfully utilizing the benefits of business analytics. Majority managers and executives of organizations using Business Analytics as an effective system of managers business transaction. However, some managers and executives consider it as a complicated system to use.

Now a days/every business organization wants to use the system of business analytics for proper functioning of business operations but distribution of analytics tools through out the organization is a very difficult task. Because, various issues and challenges may arise during the adoption of new technologies. Some of the challenges may be related to culture, people, processes and so on.

In addition to this, the major issue or challenge which may arise is the application of Business Analytics as per the business needs. The tools of Business Analytics has been

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD. ————————————

used as to identify white-color theft in organizations. These can be effectively used in identifying the inflated invoices, customer impersonation and similar offences.

Benefits of Business Analytics

The various benefits of Business Analytics (BA) are as follows,

1. It helps the organizations in measuring how much the mission statement is accomplished.

2. It provide accessibility to essential information data through which organizations can make smart, proper and accurate business decisions.

3. It provides clear and appropriate future insights to the organization through visualization.

4. It improves the efficiency of business organizations. As the Business Analytics is helpful in gathering large amount of information data at faster rate with visual representation, organization can formulate and design plans or decisions for achieving goals.

5. It helps in synchronizing the strategies related to financial and operational transactions.

6. It helps in increasing revenues and improving competitiveness.

7. It helps in maintaining effective collaboration and co-ordination between different strategies.

8. It helps in responding to the needs of customers on time by using available information of customer needs and requirements.

Q10. Write about the success of Business Analytics.

Answer :

The success rate of Business Analytics in the present business environment is increasing with a good speed. According to the report of IDC, the organizations which successfully implemented and utilized Business Analytics observed high returns between 17% to 2000% with a median Return on Investment (ROI) of 122%.

The main purpose of implementing Business Analytics is to identify the white-color theft in organizations. Through, business analytics, organization will be able to identify not only internal frauds but also external frauds. BI helps to integrate internal data with data warehouses for fraud analysis. When internal data compared with external data, the patterns and anomalies become identifiable. Moreover, suspicious activities also can be isolated, measured and tracked through fraud analysis.

Example

Williams-Sonoma corporation is an example which achieve success through business analytic techniques. It save millions of rupees with targeted marketing and multi-channel branding by using SAS data mining software.

The success of Business Analytics in recent years is summarized in the following points,

1. The usage of Business Intelligence and Analytics is relatively high and growing.

2. Large organizations are using BA more consistently compare to small organizations.

3. In 2002, most of the successful companies invested 50% income on Business Intelligence Technologies compare to Un-successful companies.

4. Government organization more frequently uses Business Analytics tools compare to any other organization or sector of economy.

5. Few companies are happy their competitive intelligence practices.

6. The companies which have used Business Analytics tools are confident on the information which they have gathered regarding customers.

4.2 DATA VISUALIZATION

4.2.1 Definition, New Direction in Data Visualization

Q11. Write in short about data visualization and visualization spreadsheets.

Answer :   *Model Paper-I, Q9(b)*

Data Visualization

Data visualization is the concept that refers to the technologies supporting visualization and data interpretation at various instances in data processing. The tools of it enable the user to identify the relationships like trends. Visualization involves digital image, GIS, graphical user interfaces, virtual reality, graphs, videos, animation and dimensional presentations.

Data visualization allows BA through the use of web based tools. Manager can view the vital organizational performance data through browser interface in real time. This would be better approach than waiting for reports or comparing the number columns. The problems which are not identified by standard analysis methods will be detected by the engineers, managers or other professionals through visual analysis technologies.

Data visualization can be implemented easily if the data is maintained in data warehouse or multidimensional special database or server. Its environment represents the data with which decision makers can view the depth and worth in real time.

Visualization Spreadsheets

Spreadsheets are the tools used by the end user's for programming decision support applications. Microsoft Excel provides general mathematical querying, report generation, statistical and other BI tools. It is a powerful and easy to use tool that is used for data manipulation. Other than this it is also used for collecting, analyzing and summarizing the data from several sources. Spreadsheets can be applied in

tracking travel expenses, forecasts, creating reports, devising budgets etc. The programmers are capable of creating advanced pivot tables and macros with which work of programmers get emulated. Excel can be leveraged through visualization in several ways such as by focusing on communications, improving the effectiveness, providing comprehension, powering collaboration, popping out a anomalies etc. Examples of it are Risk management and understanding MSS models.

Q12. Discuss in brief about new directions in data visualization.

Answer :

Data visualization is applied to mainstream computing and intelligent visualization. The main stream computing integrates the data visualization with decision support tools and applications. The intelligent visualization performs data interpretation. Along with decision support tools, most of the OLAP vendors provide three-dimensional visualization tools also. Consider for example Forest Tree 6.0 that is a web enabled development tool containing three-dimensional visualization version that allows the users to visualize as well as manage several dimensions in single view. Certain new visual tools are being developed in order to analyze web site performance. An example of it is ADVIZOR solutions.

Dashboards and Scorecards

The executives who are busy mostly require data visualization. Charts, tables and graphs are loaded into the EIS of 1990's. Later on they evolved into management cockpit products, dashboards and scorecards.

Visual Analysis

The nontechnical users who are gathering information from business data can perform enterprise data analysis. An example of it is visual, a database query language which empowers Hyperion's visual explorer. Most of the companies provide tools for visual analysis that can be performed interactively.

Financial Data Visualization

The BI data visualization is applied even in the area of finance. The Chief Financial Officers (CFO's) need an assurance that the processing power of computer is tempered with human being insight. This is required to prevent the systems from identifying the patterns that are meaningless in data. Data visualization can be used for performing such task. For example a program is developed through which CFO can fly on three-dimensional land scape. This will represent risk, return and liquidity of company's assets. Laster-tree dimensional visualization is used in cyberspace.

4.2.2 GIS, GIS Vs GPS

Q13. Explain about GIS.

Answer :   *Model Paper-II, Q9(b)*

Geographical Information System (GIS) is a computer based system which enables us to capture, store, analyse and display the data geographically from a base location.

GIS is an advanced technology, which can control the process of locating and attributing the data of various spatial features.

The development of GIS took place in the year 1960. In the late 1970's people preferred to use computers for storing, processing and analyzing geo-spatial data. But due to its complicity, many experts and eminents observed difficulty in understanding and analysing the spatial data. Then later in the year 1990 with the help of Graphical User Interface (GUI) the working of GIS became easier.

GUI is affordable tool for software and hardware and it increased the range of applications of GIS. Now a days, the application of GIS is used widely in many fields such as water resources, forestry, agriculture, land use, land cover etc., which includes extensive functionality for processing of image data.

A comprehensive geographical information system requires a method for,

1.   Gathering data input from various sources like maps, aerial photos, satellites.

2.   Performing data storage, retrieval and query process.

3.   Performing data transformation, analysis and modeling.

4.   Delivering the data in the form of maps, reports and plans.

In other words the GIS can be defined as, a system consisting of hardware, software and procedures that are intended to support modelling, capturing, checking, restoring, integrating, manipulating examining and displaying the geographically referenced data with the help of digitalized maps. These maps solve the complex problems as well as management problems. One of the characteristic features of GIS is that, each digital object has a recognized geographical location. If the user wants to generate information for planning, problem solving, decision making then he/she can combine maps with spatially-oriented databases (geographical location) and also with many other databases. As a result the productivity and standard of the decisions increases.

Generally, GIS is a dual database system, which is capable of performing spatially referred data and executing certain group of operations on the data in order to proceed with that data. There are numerous application areas that implemented GIS some of these include banking, retailing, NASA, transportation, natural resource management, agriculture, urban planning, public administration, emergency preparedness and the military.

Applications of GIS

1. GIS supplies huge amount of efficient information that is helpful in analyzing and making decisions.

2. The graphical representation of GIS helps the managers in visualizing the data without any difficulty.

3. GIS is used to enhance the decision-making process in both public and private sectors.

4. It is used in local government for mapping as well as in decision making applications.

5. It helps the organization in gaining competitive advantage by providing the organization with the information about the areas where marketing is growing rapidly.

6. In banks, GIS is used to support activities like finding out the branch and ATM location. Determining customer's residence address, age, income level, account number.

In addition to the above applications, certain examples of successful GIS applications are as follows.

	Organization	Application
1.	Sun Micro Systems	It manages leased properties at various places through entire world.
2.	Toyota and other automobile manufacturers.	It makes use of combination of GIS and GPS as navigation tool for direct the drivers to their destinations using best routes.
3.	Western Auto	It combines the data with GIS inorder to create a demographic profile of a neighbour of any store in detail and generate best product mix for the store.
4.	Pepsi, Cola, Inc., Super Value, Acodia, Inc.	It makes use of GIS for selecting the sites of new Taco Bell and Pizza hut restaurants.
5.	Sears, Roebuck & Co/Kmart	It uses GIS for supporting truck routes planning.
6.	Wood Personnel Services	It maps the neighbours where temporary workers stay for knowing the location of marketing and recruiting cities.
7.	CellularOne Corporation	It maps the complete cellular network for identifying the clusters to call disconnects. It dispatches the technicians also.
8.	Consolidated Rail Corporation.	It monitors the condition of railroad track upto 20,000 miles and parcels of adjoining land.

Q14. Write about GIS Vs GPS.

Answer :   *Model Paper-III, Q9(b)*

Geographical Information System (GIS)

For answer refer Unit-IV, Page No. 4.8, Q.No. 13.

Global Positioning System (GPS)

Global positioning system is a space-based weather radio navigation system by which the time, location and velocity of an object can be determined any where on the globe at any time. With the help of proper receiving equipment this system allows exact, continuous, three-dimensional location and velocity information to the user. It uses the satellite signals to produce distances in order to triangulate location anywhere on earth.

Global positioning system consists of three segments, as listed below,

1. Space segment
2. Operational control segment
3. User equipment.

1. Space Segment

It is a satellite constellation segment in which user equipment receives data messages and ranging signals through satellites present in orbit.

2. Operational Control Segment (OCS)

Operational control segment detect and maintain the orbital configuration of the satellite in space. It reviews the satellite clock corrections and other essential parameters required for finding the location of user, velocity and time.

3. User Equipment Segment

User equipment receives the L-band signals transmitted from the satellite and find the location of user, velocity and time.

This global position system determine the position of moving object with the help of various survey techniques.

SHORT QUESTIONS AND ANSWERS

Q1. Define Business Analytics write any two pre-requisites for effective business anlaytics.

Answer :

Business Analytics

Business Analytics refers to the broad category of skills, tools and techniques employed for gathering, storing and analysing of business related data and information by any business enterprise. The main aim of business analytics is to help the management to arrive at an objective and accurate business decisions.

Prerequisites for Effective Business Analysis

The prerequisites for effective business analysis are as follows,

(i) Top Management Support

It is necessary for the full support of the top management for the effective and result oriented implementation of BA.

(ii) Team Work

Business analytics team need to work in sync with other departments of the organizations for collecting the required data and information which would be used for preparing BA reports.

Q2. Explain briefly why Business Intelligence/Business Analytics projects fail.

Answer : *Model Paper-III, Q4*

Although BI and BA hold immense potential to make the organization prosperous and successful, sometimes BI projects fail. The following are the major reasons for failure of B.I/B.A projects,

1. Lack of top management support.
2. Failure to consider BI projects as a cross functional business activity.
3. Lack of skilled staff to effectively manage BA/B.I projects.
4. Not following a sequential breakdown of BI activities.
5. Resource crunch with regard to functional department manpower or a cute financial constraints.

Q3. Define Data visualization.

Answer : *Model Paper-I, Q4*

Data visualization is the concept that refers to the technologies supporting visualization and data interpretation at various instances in data processing. The tools of it enable the user to identify the relationships like trends. Visualization involves digital image, GIS, graphical user interfaces, virtual reality, graphs, videos, animation and dimensional presentations.

Data visualization allows BA through the use of web based tools. Manager can view the vital organizational performance data through browser interface in real time. This would be better approach than waiting for reports or comparing the number columns. The problems which are not identified by standard analysis methods will be detected by the engineers, managers or other professionals through visual analysis technologies.

Q4. Write in brief about visualization spreadsheets.

Answer :

Spreadsheets are the tools used by the end user's for programming decision support applications. Microsoft Excel provides general mathematical querying, report generation, statistical and other BI tools. It is a powerful and easy to use tool that is used for data manipulation. Other than this it is also used for collecting, analyzing and summarizing the data from several sources. Spreadsheets can be applied in tracking travel expenses, forecasts, creating reports, devising budgets etc. The programmers are capable of creating advanced pivot tables and macros with which work of programmers get emulated. Excel can be leveraged through visualization in several ways such as by focusing on communications, improving the effectiveness, providing comprehension, powering collaboration, popping out a anomalies etc. Examples of it are Risk management and understanding MSS models.

Q5. Define GIS.

Answer :

Geographical Information System (GIS) is a computer based system which enables us to capture, store, analyse and display the data geographically from a base location. GIS is an advanced technology, which can control the process of locating and attributing the data of various spatial features.

Generally, GIS is a dual database system, which is capable of performing spatially referred data and executing certain group of operations on the data in order to proceed with that data. There are numerous application areas that implemented GIS some of these include banking, retailing, NASA, transportation, natural resource management, agriculture, urban planning, public administration, emergency preparedness and the military.

Q6. What is GPS?

Answer :   *Model Paper-II, Q4*

Global positioning system is a space-based weather radio navigation system by which the time, location and velocity of an object can be determined any where on the globe at any time. With the help of proper receiving equipment this system allows exact, continuous, three-dimensional location and velocity information to the user. It uses the satellite signals to produce distances in order to triangulate location anywhere on earth.

Global positioning system consists of three segments, as listed below,

1. Space segment

2. Operational control segment

3. User equipment.

INTERNAL ASSESSMENT

I. **Multiple Choice**

1. The subject of business analysis revolves around the concept of _____.   [   ]

    (a) Decision making

    (b) Strategy decision

    (c) Web based process

    (d) None of the above

2. _____ can be defined as the art and science of analysing any data or information.   [   ]

    (a) Business intelligence

    (b) Data mining

    (c) Visualizations tools

    (d) Analytics

3. The concept of balanced scorecard was primarily developed by _____.   [   ]

    (a) Kaplan and Norton

    (b) B.A Professions

    (c) Oracle Inc

    (d) None of the above

4. _____ refers to the click of the internet pages and provides complete information of the websites.

    [   ]

    (a) Business analytics

    (b) Click stream analysis

    (c) Business intelligence

    (d) Data mining

5. Some of the tools/products offered by vendors for advanced business analytics are _____.   [   ]

    (a) Microsoft incorporation

    (b) Cognos 8

    (c) ILOG

    (d) All of the above

6. _____ are tools used by end users for programming decision support applications.   [   ]

    (a) GIS

    (b) GPS

    (c) Data visualization

    (d) Visualization spreadsheets

7.   Three-dimensional visualization is used in _____.   [   ]

     (a)   Marketing

     (b)   Finance

     (c)   Cyberspace

     (d)   None of the above

8.   Trends can be identified by data visualization _____.   [   ]

     (a)   Methods

     (b)   Tools

     (c)   Techniques

     (d)   Types

9.   Web site performance can be analyzed with _____.   [   ]

     (a)   Visualization

     (b)   GIS

     (c)   GPS

     (d)   Visual tools

10. _____ delivers data in the form of maps, reports and plans.   [   ]

     (a)   GIS

     (b)   GPS

     (c)   Visualization

     (d)   Spreadsheets

II.   Fill in the Blanks

1.   _____ refer to integration and applicability of business analysis activities with web based processes.

2   _____ is the most important prerequisite for business analytics to be effective.

3.   Business analytics is known by multiple other names such as _____, _____ and _____.

4.   _____ refers to different variety of activities which are performed by the end uses over the internet.

5.   _____ tools look similar to OLAP.

6.   _____ allows BA through the use of web based tools.

7.   Data visualization environment allows decision makers to view _____ and _____.

8.   _____ allows to store, analyze and display the data geographically from base location.

9.   GIS is a _____.

10. _____ is a space based weather radio navigation system.

KEY

I. **Multiple Choice**

1. (a)
2. (d)
3. (a)
4. (b)
5. (d)
6. (d)
7. (c)
8. (b)
9. (d)
10. (a)

II. **Fill in the Blanks**

1. Web analytics
2. Goal/mission statement
3. Business intelligence, BI application and BA processing
4. Online Analytical Processing (OLAP)
5. Data mining
6. Data visualization
7. Depth, Worth
8. GIS
9. Dual databox system
10. GPS

III. Very Short Question and Answers

Q1. Define Business Anlaytics (BA).

Answer :

Business Analytics (BA) refers to the broad category of skills, tools and techniques employed for gathering, storing and analysing of business related data and information by any business enterprise.

Q2. What is Web Anlaytics?

Answer :

Web analytics refers to the integration and applicability of business analytics activities with web based processes. In simple terms web analysis refers to the interface between BA and the Internet.

Q3. Write a short note on click Stream Analysis.

Answer :

Click stream analysis refers to the analysis or study of the mouse clicks of the internet users computer. Further, it also includes the recording, storing and detailed analysis of the sequence in which the web user had visited the various web pages.

Q4. Define Data visualization.

Answer :

Data visualization is the concept that refers to the technologies supporting visualization and data interpretation at various instances in data processing.

Q5. What is GIS?

Answer :

Geographical Information System (GIS) is a computer based system which enables us to capture, store, analyse and display the data geographically from a base location.

Q6. What is GPS?

Answer :

Global positioning system is a space-based weather radio navigation system by which the time, location and velocity of an object can be determined any where on the globe at any time.

UNIT 5

Business Intelligence Implementation

After studying this unit, one would be able to understand,

- ❖ Implementation of Business Intelligence.
- ❖ Business intelligence implemental factors and critical success factors.
- ❖ Managerial issues of BI.
- ❖ BI and integration implementation.
- ❖ Growing trends in business intelligence.
- ❖ Social networks and business intelligence.
- ❖ Reality mining.

INTRODUCTION

The Business Intelligence (BI) holds strategies and technologies generally implemented by enterprises pertaining to data analysis of business information. Some of the most prevalent business intelligence technologies could be reporting, online analytical processing, analytics, data-mining, process mining.

In actuality, the Business Intelligence (BI) can be defined as a set of techniques or tools used for acquisition and transformation of raw-data into information for business analysis purposes.

It also discusses the issues related to system development and the need for integration, cost-benefit issues, legal issues etc. Further more, it gives deep understanding of types of integration, levels of BI integration and embedded intelligent systems the emerging trends in Business Intelligence includes social networks and Business Intelligence, RFID and New BI Application Opportunities, Reality Mining and Collaborative Decision Making.

5.1 IMPLEMENTING BUSINESS INTELLIGENCE

5.1.1 Implemental Factors, Critical Success Factors of Business Intelligence Implementation

Q1. What are the factors that affect the implementation of business intelligence. Also, list its critical success factors.

Answer :   *Model Paper-I, Q10(a)*

BI Implementation Factors

There are various factors that effect BI implementation such as technological, administrative, behavioral etc. Most of these factors are generic to many information systems and have been widely analysed in information system literature. Some of the crucial factors that influences the decision making process of BI are as follows,

(i)   Reporting and Analysis Tools

The following are the details which contributes in the poor decision making process, features and functionality, scalability and deployability, usability and manageability, modifying the applications.

(ii) Database

The following are the details which contributes in the poor a decision making process, scalability and performance of database, security and customization, manageability and availability, write back operation.

(iii) Extraction, Transformation and Load (ETL) Tools

The following are the details which contributes in poor decision making process, capability of reading any source, efficiency and productivity, cross platform support.

(iv) Costs Involved

The following are the details which contributes in poor decision making process, hardware cost, software cost, internal and external development cost, maintenance cost etc.

(v)   Benefits

The benefits of BI implementation are as follows,

1.   BI implementation saves the time and improves the operational efficiency.

2.   It incurs less operational cost.

3.   It enhances the various services of the customers and their satisfaction.

4.   It enhances operational decision making as well as strategic decision making.

5.   It enhances knowledge sharing (i.e.,) the exchange of knowledge among research teams.

6.   It enhance the interaction between the employees and provides a satisfactory conclusion of all employees.

Critical Success Factors of Business Intelligence Implementation

The critical success factors of BI implementation are as follows,

1.   Driving methodologies based on business and managing projects.

2.   Focusing on the vision and planning.

3.   Maintaining management and sponsorship.

4.   Solving the problems related to data management and maintaining its quality.

5.   Providing solutions based on the requirements of the users.

6.   Considering the performance of BI system.

7.   Providing powerful and extensible framework.

5.1.2 Managerial Issues Related to BI Implementation

Q2.   Discuss in detail about the managerial issues related to BI implementation.

Answer :   *Model Paper-II, Q10(a)*

The following are the managerial issues in regard to BI implementation,

1.   System Development and the Need for Integration

An effective BI application is difficult to develope. In account of this, the dealers provides complex integrated application groups which incorporates enterprise resource planning (ERP) and customer relationship management (CRM). The important dealers of BI are Oracle, Microsoft, IBM etc., offers application integration that are web enabled.

2.   Cost-benefit Issues and Justification

An effective BI solution is expensive and is substantiated only in large organization. On the other hand, such situation can put negative effects on the small organization, which gambits (turn the situation into an advantage) by making solutions cost effective. This is purely done by influencing present databases and not creating additional database. Subsequently, on-demand BI comes to rescue in such situations.

3.   Legal Issues and Privacy

An effective BI recommends companies to issue electronic or printed catalogs or promotions to a certain age group or gender.

In certain case, a man used a high end brand for offering large amount of discounts to other person based on the gender. This was considered an encroachment of individual privacy.

4.   BI and BPM Today and Tomorrow

An effective business information quality and timelines pertaining to an organization is not taken into account as an option between profit or loss rather it is a matter of continuity. According to the recent research, it is studied that BPM dashboards and business analytics (BA) will be predominantly used by many people. This produced huge benefits as BI got extended to many uses. Thus adding value to the enterprise. Thus, maximizing the utilization of present data assets. Thus, BI makes use of the present IT technologies to influence their IT investments and utilize their legacy and real time data.

5.   Cost Justification: Intangible Benefits

An effective enterprise systems offers tangible benefits whereas the intangible benefits such as mortage crisis, political unrest etc should be clearly justified.

6.   Documenting and Securing Support Systems

An effective decision support or BI modules can be developed by the employees for increasing the productivity and the quality of the work.

A list of the stock productivity is recommended to maintain. This is because if an employee is on leave, the productivity tool remain same. Furthermore, security measures must be properly done. This is necessary because, the developers of BI application are not professionals due to which data integrity and security of the system gets compromise.

7.   Ethical Issues

An effective BI and predictive analysis can produce ethical issues such as privacy and accountability. The mistake can not only harm other employees work but also the whole organization.

For instance, Decision Support System (DSS) is developed by an organization for assisting the employees in computing financial implications for early retirements of employees. But, if the developer forgets to include tax implications then the entire retirement statement becomes incorrect.

8.   BI Project Failures

An effective BI project is likely to be fail due to the human mistakes and software failures.

The following are the examples of BI project failure,

(i)   BI projects cannot be recognized as an enterprise-wide business initiatives.

(ii)   Availability of less number of business sponsors.

(iii) Availability less number of educated employees.

(iv)   High dependent on sellers.

(v)   Lack of coordination among the business representatives.

5.1.3 Business Intelligence and Integration Implementation – Types, Need, Levels of Business Intelligence Integration

Q3.   What is the need of BI software integration?

Answer :

The following are the objectives of BI software integration.

1.   Implementing BI

As BI systems generally function by connecting to data sources, utilities, other applications etc., it is necessary that the connection between them should be efficient and effective.

2.   Increasing the Capabilities of BI Applications

Every BI development tool performs a specific task individually. This individual functioning of several BI development tools enable them in enhancing one another.

Example

When Business Analytics (BA) can be used for suggesting the plan of optimal resource allocation. On the other hand, an attached dashboard can be used for notifying the management about variations from their plan.

3.   Enabling Real-time Decision Support

In a real-time environment, decision making is supported by performing tight integration. An example of such real time decision support is the use of wireless communication and web services in the transportation system for promoting the data flow.

4.   Permitting More Powerful Applications

It implements more powerful applications such as intelligent system in order to give real time capabilities.

5.   Helps in System Development

It enables rapid development of applications and interaction between the system components by employing tighter integration.

6.   Enhancing Support Activities

It enhances the performance of BI applications by providing various support activities.

Q4.   List and explain different types of integration. Also, discuss different levels of BI integration.

Answer :

Types of Integration

When integration is performed on computer based systems each individual component of a system operates as a single entity. Integration can be performed either at development level or at application system. It can be classified into two categories,

(i)   Functional Integration

(ii)   Physical Integration.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD. ——————————————

(i)   Functional Integration

Functional integration refers to a process where group of distinct applications are supported by a single system.

For instance, different functions like using an email, a spreadsheet, interacting with external databases, data storage and data manipulation can be performed by using a single workstation.

Functional application integration can be performed by using the following two methods,

(a)   By integrating more than one decision support application for developing unified application.

(b)   By integrating more than one BI tools with another information systems like knowledge management, databases or financial system.

(ii)   Physical Integration

Physical integration refers to the process of combining hardware, software and other characteristics of communication that are required for achieving functional integration.

Levels of BI Integration

Functional integration can be performed at two different levels (i.e.,) across BI and within BI. These levels are not only suitable for solving the repetitive or sequential decision problems of a system but also can be used for transferring the output of one system as input to the other system. Integration across BI can be seen by collaborating multiple analytics wherein every single analytics specifies a particular part of a complex decision problem.

The next level of integration (i.e., within BI system) refers to the process of collaborating various BI technologies while developing a complex BI system.

5.2 EMERGING TRENDS IN BUSINESS INTELLIGENCE IMPLEMENTATION

Q5.   What are the developments predicted by Gartner in BI market?

Answer :

The following are the Gartner's predictions for the development of BI market,

1.   Minimum 40% of the entire budget will be managed by business units by the year 2012.

2.   20% of the companies would employ industry specific analytic applications as most reliable component of their business intelligence portfolio untill the year 2010.

3.   Collaborative decision making would surge(rise) as a new product by collaborating social software and business intelligence platform by the year 2009.

4.   20% of the analytic application corresponding to business processes might be transferred by using coarse grained application mashups by the year 2012.

5.   By the year 2012 about 35% of top companies might not succeed in making an appropriate decisions on the changes existing in business and markets due to insufficient data, processes and tools.

Q6.   Discuss in detail about Web 2.0 as an emerging trends in Business Intelligence.

Answer :   *Model Paper-III, Q10(a)*

Web 2.0 Revolution

Web 2.0 adopts an architecture of participation which inspires not only user's communication but also community contributions. On the other hand, web 1.0 emphasizes only on small companies and advertisers generating content which can be helpful for users to access. Almost all the web 2.0 companies depends on the content created by users. It is not important how much content is generated by the users but it is important where and how that content is being used. Google, the most popular search engine and advertising company helps users in publishing their content on the website. Google is providing a platform to the users. There are many other sites that provide platform to the users such as flicker, My space, wikipedia and youtube.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

Architecture of participation also plays a significant role in developing softwares. These type of softwares are called open source softwares which are freely available, and can be modified as well as used without implying rules. These softwares developed by users are far better than the softwares developed by software developers. There are many popular social networking sites like facebook, Bebo, Linkedin and avatar a 3d virtual world which helps in communicating with people around the world personally, whereas social networking sites allows their users to share, tag the content as well as recommend their favourite sites.

Representative Characteristics of Web 2.0

1. Web 2.0 becomes more popular and valuable when the number of users increases and its ability in gaining the knowledge or ideas of the users also increases.

2. It can be mashed up more oftenly via web service interfaces.

3. It depends on the user generated as well as user controlled data.

4. It provides various light weight programming techniques and tools which enables any user to operate as a web site developer.

5. It permits users to use any application completely.

6. It mostly depends on the social networking, social computing, social software etc.

7. It quickly develops the models of new business and provides modern support to the information sharing and collaboration.

8. By using web 2.0 as a platform for application development the virtual elimination of software upgrade cycles permits rapid prototyping and maintains the work in progress.

Q7. Discuss in detail about virtual worlds.

Answer :

Virtual World

Virtual world is a technology that creates an artificial environment that responds to, and is controlled by the behavior of the user. It is created by using interactive graphics software which is so closed to reality that people feel this artificial reality really does exist. The same computer-generated artificial environment can be shared by more than one person and even a large group. In many virtual reality systems, the user wears stereo goggles and headset to see and hear the environment. It wears a computerized display and gloves that contains hand position sensors to immersed in the virtual system. All these equipments accept and transmit the real time information back to the computer.

Virtual world has many applications in education, entertainment, scientific and business work. For example, Virtual world helps surgeons to reduce bleeding and trauma during brain tumors surgery. The three-dimensional modeling also helps financial decision makers in analyzing the data in a better way by using visual, spatial and aural immersion, virtual systems. For example, they can visualize the stock futures with different colors, hue and intensity which indicates the deviation from current share prices. Other information such as current trends or debt/equity ratio can be conveyed using sound effects.

Multimedia is applicable even in virtual reality technology. This technology enables a user to perform interaction with a computer simulated environment irrespective of its physical presence. This interaction is done using standard input devices (such as keyboard or mouse) or through multimodal devices (such as wired glove or omnidirectional trade mills).

Second life holds about millions of users. It which allows them to create Avatars indicating their own identity which can help them in searching users of same interest. It is developed by Linden Labs as a 3D virtual world. Second life has helped many people in conducting business and earning profits. For example, lawyers create their avatar and meet clients. Most popular companies like IBM and Hewlett-Packard makes use of second life to communicate with customers, conduct meetings and hire new employees.

Facebook is the most widely used social networking website having millions of users. It was found by Mark Zucker berg, Dustin Moskovitz and Chris Hugles at Harvard. Initially, it was limited only to the students of Harvard but later on, it was extended to each and every individual above 12 years of age. The features that makes it widely adopted include uninterrupted data adding facility with scalable high performance computing features.

It is a booming site in today's world where over 185 million users can interact with each other, add friends and interchange messages. Users can also update their personal profiles so as to notify their friends about themselves. All these capabilities are provided free of cost and without restrictions.

The network associated with facebook employs two data centers that minimizes the cost associated with its management. Certain open source technologies are also included for a smoother experience. Uses of cloud computing offers an integrated platform with which, applications from third-party can easily provide their services. This integration is done by using application programming interfaces (APIs).

Cloud services are provided by using multiple technologies such as MySQL, PHP, Linux and Apache that collaborately handle operations like newsfeeds, notifications etc. While serving queries generated from users such as search, most relevance results are generated with use of social graph of the user. This graph is composed of various user related data. Moreover, for a faster response of such queries, a cluster of MySQL instances is used.

Developers employs certain inbuilt tools such as thrift. The thrift tool allows integration of various services developed in different languages to work together efficiently. This feature is provided with use of a set of interfaces capable of allowing cross-language development

5.2.1 Social Networks and Business Intelligence

Q8. Define social network. What is the size of social network sites? Discuss about social network analysis software.

Answer :

Social Network

Social Networks are the networks that allow the users to interact with each other through messages, comments, images, posts etc. It is incorporated with social interactions as well as personal relationships.

Size of Social Network Sites

The size of the social networking sites depends upon the number of users that joins the website per year. The speculated annual growth recorded in first year was 40-50% and in next few years, the growth rate diminished by 15-25% per year.

Social Network Analysis Software

It basically provides quantitative and qualitative analysis of social networks. This software extrapolate the network features using numerical or visual representation. The network can ranges from project teams, classrooms, sports teams, legislatures, nation-states etc. They forms direct-connections between nodes and depending upon shared attributes and attendance forms indirect connections. It examines relational data and caters the statistical routines and visualization mechanisms. In simple terms, the software identifies, represents, analysis, visualize and simulates the network nodes (which could be agents, organizations or knowledge) and including edges (relationships). These operations are carried out from multiple types of input data which also includes mathematical models of social network.

Apart from this the availability of network analysis tools eased the released of new products in different forms and sizes. One such type of famous representation is visual representation that is used for understanding of network data and for displaying the analysis results.

The representative tools which substantiates the above presentation are as follows,

1. Business oriented social network tools like In Flow, Net Miner etc.

2. Social Networks Visualizer (SocNetv) is an open source package in Linux.

Q9. Discuss in detail about mobile social networking.

Answer :

Mobile Social Networking

The mobile social networking is a social networking which allows the users with common interest to collaborate converse using mobile phones or tablets. In other words, it gives new ideas of communicating with people using internet. In a way similar to web-based social networks, here also virtual communities are implemented. It implements mobile messaging applications which is one of the efficient method for offering smoother user interaction. The user check mails, sent messages and post content on the web using their cell phones and PDA's easily without any hastiness. The newly introduced I-phones releases the dream of using internet fully in phones but not the simplified version which is embedded in mobile phones.

The two types of mobile social networks are as follows,

(i) First type incorporates those companies which uses a default cell phone browser with a wireless carrier in order to distribute the communities.

Example

Myspace, using AT & TS wireless network.

(ii) Second type incorporates those companies which does not maintain carrier relationship among the communities. They adopt other methods to attract customers.

Further more, the user can see windows live space on a small screen with slow data connection on the mobile phones. It offers social features such as profile updation photo sharing, blog entries and web etc. Moreover, this application develops many other features for improving the experience of the user with handled devices.

Mobile social networking is widely used in the countries like Japan, China, South Korea etc., because of their good mobile network facility and cost-effective data pricing. With the advent of mobile social networking, the users can access the internet anytime through mobile phone and also provides a good facility for the users who cannot easily or regularly access the system.

In today's highly competitive world, the current social networking softwares are just not restricted to exchanging of text messages but also include many advance features such as sophisticated interactions which takes place between the internet virtual communities.

Mobile Enterprise Networks

The mobile enterprise can be defined as a corporation or large organization which carries out various functionalities related to critical business. It allows employees to make use of mobile devices to access the mail, managing projects and documents. It also offers customer relationship management, carries out enterprise resource planning etc. And also many of the organizations developed their mobile-based social network.

Example

In the year 2007, a coca-cola company developed a mobile-based social network which can be accessed only via a mobile phone. This is used by the company order to attract the users to use the product.

Mobile Community Activities

In mobile social networks, the users can creates their personal profile, make friends, join in the chat rooms, develop a chat room etc., via a mobile phone. The users can also send photos, videos, documents etc., to other users. However, some of the companies permit the users to develop their own mobile community.

Q10. Give an overview of any two social networking services.

Answer :

The two most popular social network services are,

1. Facebook

2. MySpace.

1. Facebook

Facebook is the most widely used social networking website having millions of users. It was found by Mark Zucker berg, Dustin Moskovitz and Chris Hugles at Harvard. Initially, it was limited only to the students of Harvard but later on, it was extended to each and every individual above 12 years of age. The features that makes it widely adopted include uninterrupted data adding facility with scalable high performance computing features.

It is a booming site in today's world where over 185 million users can interact with each other, add friends and interchange messages. Users can also update their personal profiles so as to notify their friends about themselves. All these capabilities are provided free of cost and without restrictions.

The network associated with facebook employs two data centers that minimizes the cost associated with its management. Certain open source technologies are also included for a smoother experience. Uses of cloud computing offers an integrated platform with which, applications from third-party can easily provide their services. This integration is done by using application programming interfaces (APIs).

Cloud services are provided by using multiple technologies such as MySQL, PHP, Linux and Apache that collaborately handle operations like newsfeeds, notifications etc. While serving queries generated from users such as search, most relevance results are generated with use of social graph of the user. This graph is composed of various user related data. Moreover, for a faster response of such queries, a cluster of MySQL instances is used.

Developers employs certain inbuilt tools such as thrift. The thrift tool allows integration of various services developed in different languages to work together efficiently. This feature is provided with use of a set of interfaces capable of allowing cross-language development.

2. My Space

According to Hitwise report, My Space is at the top of social networking sites used by distinct users such as friends families etc. Each user have their own page in which they can create their profile with general information. They can even add media files and tag files to other friends. They can even chat in the provided personal chat box.

Myspace has become popular among youngsters as it is the most popular networking site which allows free advertisements, news corp etc.

Q11. Write about representative areas and examples of enterprises social networking.

Answer :   *Model Paper-I, Q10(b)*

Finding and Recruiting Workers

E-Recruitment

In today's highly competitive world, most for the companies are using internet for recruiting talented aspirants. Recruitment for positions from job seekers through internet is called as "on-line recruitment" or "e-recruitment". In e-recruitment, job seekers post their resumes on job sites such as Monster.com, Naukri.com etc., whose accession is free for both domestic companies as well as for MNCs. Companies access these sites for recruitment and selection of potential candidates which are as per their requirements. Using e-recruitment, an organization can recruit job seekers belonging to any part of the world.

The various methods of e-recruitment which are used by organization are as follows,

1. Job

2. Agency

3. Professional/career websites

4. Company's own website.

Employment Placement

Job market is uncertain and has grown rapidly in the upcoming years. Increased number of people are emphasizing on online employment placement programs where job seekers, recruiters meet, instead of opting the conventional job markets methods such as commercial employment agencies, head hunting companies, corporate recruiters, newspaper ads. Traditional job markets has some drawbacks such as increased cost, limited information, receipt of applications after the deadline, loss of materials from the applicants.

Online Job Market

Online job market is being used by thousands of users such as recruiters, job seekers, recruiting agencies. It has become popular because of its effective use over the internet. Job seekers can apply for jobs by visiting online job websites and by registering their details like skills, personal information and the kind of job they are interested in joining. On the other hand even recruiters visit these sites and register themselves and search for employees who has the capability of performing the job in their organization. Some of the most popular online job markets are www.Monster.com, www.hotjobs.com, www.careermosaic.com/cm.

Management Activities and Support

The applications in management activities and support are generally associated with managerial decision making which is basically done depending on the analysis of collected data in social networks. The common examples of it includes determining key performers, locating experts, determining right candidates, providing innovative ideas and solutions to complicated issues.

In training session, multiple companies are using enterprise social networking and virtual world.

Example

A company such as Cisco tries to use the virtual teaming campus in second life.

Knowledge Management and Expert Location

In this category, it incorporates application activities like knowledge discovery, maintaining, sharing, transferring, creating and disseminating the applications.

Example

Consider a social networking site such as innovative.com where in large number of people are participating including scientists inorder to solve the problems related to science.

Enhancing Collaboration

In social networking collaboration can be performed internally as well as externally. Internal collaboration is performed among employees from distinct streams.

Example

Virtual teams

Where as external collaboration involves working with customers, suppliers and some other business partners. Besides this, it can be used frequently in forums different types of groups and by sing wikis and blogs.

Using Blogs and Wikis within the Enterprise

According to the reports of Jefferies 2008, studies revealed that 71% of the top companies use blogs while only 64% of them use wikis. So as to perform the following applications,

(i)   It is used in project collaboration and communication.

(ii)   It is used in Frequently Asked Questions (FAQs).

(iii)   It helps in Electronic learning and training upto 46%.

(iv)   It facilitates about 63% in process and procedure document.

(v)   It is used upto 41% forums for providing innovative ideas.

(vi)   It supports upto 24% in collaboration with customers.

5.2.2 Collaborative Decision Making

Q12. What is meant by collaborative decision making? List its benefits.

Answer :

Collaborative decision making refers to a process where group of people in an organization collaborate on a specific data and discuss about its interpretation. The analysis of Gartner Inc researchers revealed that CDM drastically improves the decision making quality by integrating the data in BI systems with collaborative input. The following are the key benefits of collaborative decision making (CDM),

(i)   It serves as a decision support system for nonroutine and complex decisions that need iterative human interactions.

(ii)   It improves decision process and the data corresponding to it by performing adhoc tagging.

(iii)   It allows organizations to define the business value of BI more effectively by combining BI with decisions and outcomes that can be evaluated.

In the earlier decades it was not easy to expect a good decision by just providing huge access to data. Instead several social, cultural and educational factors were required to determine how well organizations are capable to enhance their ability of decision making. Inorder to overcome these issues CDM was brought into existence which rectified the fault in decision making by adding the missing features.

Q13. Discuss how CDM can be used in virtual teams decision making.

Answer :

Collaborative Decision Making

The collaborative decision making (CDM) in general can be defined as a joint and associated decision making process. It imparts all details pertaining to the parties participating in the entire decision-making process. Interestingly, the virtual teams facilitates itself as flexible tool in CDM. These teams operates across the huge geographic boundaries connected using computer and communication technologies. We also observe that they represents negative influence on the decision making process due to its diversed outlook such as heterogeneity in teams and with a denouement (generating result) in teams of different sets of values, implementation of non-standard technologies for interaction. Apart from this, the lack of trust prevailing between virtual teams adds a great disadvantage.

Moreover, a completely new form of work, collaboration and decision making process was enforced by many companies due to the economy rise. According to the research, the information technology market is deficient of virtual teams collaboration. It alleviates this problem by developing systems which promotes CDM process with the help of social software. In support to this, the business application corresponding to social software techniques were first initiated by consumer driven social networks such as facebook, myspace, twitter etc. It is already implemented in the form of collaborative social software which keeps the records of where abouts of colleagues and their work. It is also used to inform the employees regarding the urgent meetings. This concepts also fulfills the task of making collaborative environments for decision makers for problem discussion, performing brain storming, focussing upon the pros and cons of the idea. The environment can be made more rich and productive by including social software elements like tagging, recommendations, rating and providing contextual information.

The quality of the decision making is improvised by combining CDM and social software with business intelligence. It makes the direct contact with the information stored in BI systems along with the input extracted from social software. Even though, Business Intelligence (BI) is primacy (dominant). It is an unstructured process which is not repetitive and lack of tools which clearly makes it disconnected from the business process. This abstains the user from viewing the business value of BI.

As we already know that, the collaboration of BI with social software may give a direct way to view the value of BI. This is due to the fact that analytical observation, understanding and details are linked to business decisions. They are also embedded in the social context.

For instance, making an investment regarding business decision, giving quotations for future revenues, expenditures and rates of interest. Once, all decisions are made, the generated results corresponding to the predictions (decisions) is compared with Key Performance Indicator (KPI). This inturn helps to guage the sales and profits in the business. Now, updation of the prediction model is done by BI platform using the authentic temporary results on the KPIs. This becomes born for KPI participants for enhancing critical threshold, rethinking of decisions.

Furthermore, another important concept could be brainstorming which helps to collaborate strategically and transform the business. It is nonroutine activity which basically, improvise discovery, innovation creation, learning and relations. There are chances that, the outcome of these manual processes can become void and lack of further formal process for decision audit, assessment can exacerbate (make worse) the entire process. Thus, CDM comes an ideal mechanism for improvising the situation.

5.2.3 RFID and Business Intelligence

Q14. Discuss in detail about RFID.

Answer :   *Model Paper-II, Q10(b)*

RFID stands for Radio Frequency Identification. RFID is an identification device used for describing several technologies which makes use of radio waves to identify people, place or objects through radio signals.

RFID technology has been practised since many years but it could not gain popularity due to its expensive installation and implementation. RFID technology is developing day-by-day and the researchers are trying to minimise the cost so that RFID can be used widely.

RFID contains a microchip inserted in a product or item in which data related to product or item is stored,

The RFID tags are two types,

1.   Active tag

2.   Passive tag

Active tags are expensive because it has in-built battery where as passive tags are inexpensive and does not have internal battery. Both active and passive tags have the capacity to read-only or read/write. These tags can be used either one time used or reusable. These tags are used to read Electronic Product Code (EPC) which is a code number used to recognize a particular item in the supply chain and also maintain record information in order to direct the work flow.

RFID tags minimise labour cost of the company on account of its good tracking system and this system is efficient.

The various application of RFID technology are,

1.   Inventory management plays an important role in retail business and this task can be performed effectively by RFID system. RFID provides alarm facility to the store keeper which helps them in maintaining appropriate stock and this results in higher customer satisfaction and huge business profits.

2.   RFID system can be used in blood banks to avoid general human mistakes. Usually, in blood banks most of the blood bottles consist of similar labels which may cause confusion to the hospital management. In order to avoid this risk, RFID tags can be attached to the blood packets which helps the medical staff in identifying correct blood group.

3.   RFID tags are used in stores/malls to avoid theft of some expensive items. RFID tags can be attached to the items or can be kept near the exist door. RFID system provide alarm facility through which one can identify the items missing in the store and can also avoid theft.

4.   RFID technology can be used in hospitals to locate patient, doctors, nurses or other staff.

5.   Apart from the above RFID applications, it is also used in animals. A small microchip is inserted into the animals skin which gives information related to animal and also its owner.

Moreover, RFID comprises 96 fits of data which is in the form of serialized global trade identification number (SGTIN). This identification number is used for recognizing cases or pallets of serialized shipping container codes (i.e., SSCC).
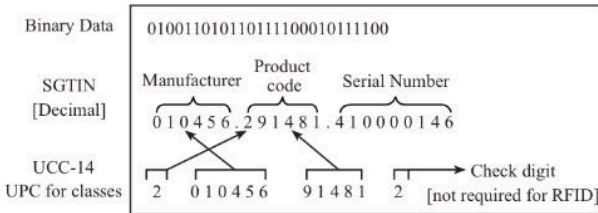
The sample of RFID Tag Data is illustrated below,

Figure: Sample of RFID Tag Data

Here, the tag is a label object which is a sequence of binary digits. These sequential binary digits are converted into its equivalent SGTIN decimal. However, SGTIN is nothing but UPC (i.e., Universal Product Code) with a difference that UPC contains the initial digit as its serial number which is used for the identification of the family of the product. However, UPC can't differentiate different cases but this drawback can be overcomed by SGTIN.

Advantages

1. RFID is a flexible identification device can be moved easily.
2. RFID tags can be read-only or read/write. The technology of RFID is versatile in nature.
3. Implementation of RFID technology results in accuracy and speed of data collection.
4. RFID tags are robust and can be used in any environments and temperatures.
5. RFID enables transmission and reception of radio signals. Since, it uses radio frequency tags and reader.
6. RFID tags can be used in stores to prevent theft, there by providing the facility of protecting expensive items in the store.
7. RFID tags have greater storage capacity.
8. RFID enables good tracking system.

Disadvantages

1. RFID technology is highly expensive.
2. If RFID tags are in liquids and metal products then, it is highly difficult to read.
3. Range of frequency differs from one country to another and it is difficult to known the working pattern of other countries where RFID tags perform their activities. It takes a lot of time.
4. At times RFID radio waves are disturbed.
5. RFID technology is considered as invasive technology.

5.2.4 Reality Mining

Q15. Discuss in detail about Reality mining.

Answer :   *Model Paper-III, Q10(b)*

Reality Mining

A new emerging technology is Reality Mining, fastly gearing up the phase which provides sense of data. It is a massive data source similar to RFID for developing data streams and for making analysis using business intelligence. This eases the process of decision making.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

Ideally, Reality mining can be specified as a collection and examination of data resulted from machine-sensed environment. This data is related to human social behavior for determining predictable patterns of behavior. The Reality mining develops a precise picture of the where about of people, keep tracks of their contacts by using Wireless devices such as mobile phones and GPS systems. Interestingly, the Big data is incorporated in Reality mining for boosting the process of research and analysis of people's interaction with technology. This effectively forms a positive transition from an individual to global community. Therefore, they advances there data analytics to a different type of data which includes real-time behavior of people working in groups which could be movement and spatial positioning. (This requires sensors).

Many businesses and customers effectively implements location-enabled devices which spreads location information through the cars, buses, taxis, mobile phones, cameras and personal navigation devices. These devices operates upon network-connected positioning technologies which are GPS, wi-fi and cell-tower triangulation. The growth corresponding to location-enabled services produces huge database related to historical and real-time streaming location information. So, subsequently keeping in mind all above, reality mining is constructed on the concept of such datasets which facilitates real-time insight with in the human-activities.

These large scale patterns serves as a basis for understanding and analyzing various classes of behaviors with in the particular context which is called as 'tribes'.

In particular, the macrosense is an application created by sense Networks. This accepts the input from all mobile devices and classifies the received data streams based on the type of customer and client. This is fulfilled once proprietary clustering algorithms are applied. Consecutively, it supports the investors in better understanding the customer patterns.

A new application called citysense is coming into picture which is adapted by sense Networks. This technology is basically used for customers in searching people who keeps similar interests. The citysense outlook can be seen in sensenetworks.com/citysense.php. The layout is in the form of black and white image displaying the patterns and roaming locations of the people in and out of the city. Also, the macrosense, which is a sense Networks's primary analytical platform examines the complete information present in the citysense for combining users and identifying tribes. It does this by performing the sampling in distribution of tribes at any point of time.

This could be further explain with the supporting idea where clubs and hip-hop clubs own their unique tribal distribution. So, the moment user checks in, the microsense initiates the selected tribe distribution from the time spent in the club. Future enhancement could be the inclusion of hotspots based on the distribution and complete activity information.

Once, the algorithms are applied, the dimensionality of the location of data gets minimized and reality mining initiates the characterization of locations based on the user activity and movement starting from huge amounts of high-dimensional location data. The algorithm unleashes the styles, meaning and relationships for delivering representation in human understandable form. This eventually makes the data suitable for intelligent predictions and also determining significant matches and similarities among places and people.

On the other hand, these technologies presents lot of problems in terms of privacy. That is, any intruder can crack the location of the cell phone. Thus, compromising the privacy of the user.

SHORT QUESTIONS AND ANSWERS

Q1. Write short notes on,

(i) Functional Integration

(ii) Physical Integration.

Answer : *Model Paper-II, Q5*

(i) Functional Integration

Functional integration refers to a process where group of distinct applications are supported by a single system.

For instance, different functions like using an email, using a spreadsheet, interacting with external databases, data storage and data manipulation can be performed by using a single workstation.

Functional application integration can be performed by using the following two methods,

(a) By integrating more than one decision support application for developing unified application.

(b) By integrating more than one BI tools with another information systems like knowledge management, databases or financial system.

(ii) Physical Integration

Physical integration refers to the process of combining hardware, software and other characteristics of communication that are required for achieving functional integration.

Q2. Write a short note on web 2.0

Answer :

Web 2.0 adopts an architecture of participation which inspires not only user's communication but also community contributions. On the other hand, web 1.0 emphasizes only on small companies and advertisers generating content which can be helpful for users to access. Almost all the web 2.0 companies depends on the content created by users. It is not important how much content is generated by the users but it is important where and how that content is being used. Google, the most popular search engine and advertising company helps users in publishing their content on the website. Google is providing a platform to the users. There are many other sites that provide platform to the users such as flicker, My space, wikipedia and youtube.

Q3. Write representative characteristics of web 2.0.

Answer :

The following are the representative characteristics of web 2.0.

1. Web 2.0 becomes more popular and valuable when the number of users increases and its ability in gaining the knowledge or ideas of the users also increases.

2. It can be mashed up more oftenly via web service interfaces.

3. It depends on the user generated as well as user controlled data.

4. It provides various light weight programming techniques and tools which enables any user to operate as a web site developer.

Q4. Write a short note on e-recruitment.

Answer :

In today's highly competitive world, most for the companies are using internet for recruiting talented aspirants. Recruitment for positions from job seekers through internet is called as "on-line recruitment" or "e-recruitment". In e-recruitment, job seekers post their resumes on job sites such as Monster.com, Naukri.com etc., whose accession is free for both domestic companies as well as for MNCs. Companies access these sites for recruitment and selection of potential candidates which are as per their requirements. Using e-recruitment, an organization can recruit job seekers belonging to any part of the world.

SIA PUBLISHERS AND DISTRIBUTORS PVT. LTD.

The various methods of e-recruitment which are used by organization are as follows,

1. Job

2. Agency

3. Professional/career websites

4. Company's own website.

Q5.   Discuss the benefits of Collaborative Decision Making (CDM).

Answer :   *Model Paper-III, Q5*

The benefits of collaborative decision making are as follows,

(i)   It serves as a decision support system for nonroutine and complex decisions that need iterative human interactions.

(ii) It improves decision process and the data corresponding to it by performing adhoc tagging.

(iii) It allows organizations to define the business value of BI more effectively by combining BI with decisions and outcomes that can be evaluated.

Q6.   List the disadvantages of RFID.

Answer :

The disadvantages of RFID are as follows,

1. RFID technology is highly expensive.

2. If RFID tags are in liquids and metal products then, it is highly difficult to read.

3. Range of frequency differs from one country to another and it is difficult to known the working pattern of other countries where RFID tags perform their activities. It takes a lot of time.

4. At times RFID radio waves are disturbed.

5. RFID technology is considered as invasive technology.

Q7.   Discuss virtual teams.

Answer :

The virtual teams facilitates itself as flexible tool in CDM. These teams operates across the huge geographic boundaries connected using computer and communication technologies. We also observe that they represents negative influence on the decision making process due to its diversed outlook such as heterogeneity in teams and with a denouement (generating result) in teams of different sets of values, implementation of non-standard technologies for interaction. Apart from this, the lack of trust prevailing between virtual teams adds a great disadvantage.

Q8.   What is Reality Mining?

Answer :   *Model Paper-I, Q5*

Ideally, Reality mining can be specified as a collection and examination of data resulted from machine-sensed environment. This data is related to human social behavior for determining predictable patterns of behavior. The Reality mining develops a precise picture of the where about of people, keep tracks of their contacts by using Wireless devices such as mobile phones and GPS systems. Interestingly, the Big data is incorporated in Reality mining for boosting the process of research and analysis of people's interaction with technology. This effectively forms a positive transition from an individual to global community.

INTERNAL ASSESSMENT

I. **Multiple Choice**

1. _____ is the booming site in today's world where over 185 million users can interact with each other and developed by Mark Zuckerberg.   [   ]

    (a) MySpace

    (b) Facebook

    (c) Twitter

    (d) Instagram

2. _____ are the networks that allow the users to interact with each other through messages, comments, images, post etc.   [   ]

    (a) Content Networks

    (b) Web 2.0

    (c) Social Networks

    (d) None of above

3. _____ adopts an architecture of participation which inspires not only user's communication but also community contribution.   [   ]

    (a) Web 0.1

    (b) Web 1.0

    (c) Web 2.0

    (d) Web 0.2

4. _____ is an identification device used for describing several technologies.   [   ]

    (a) CDB

    (b) RFID

    (c) Reality Mining

    (d) Second Life

5. The technology that enables a user to perform interaction with computer simulated environment irrespective of its physical presence is _____.   [   ]

    (a) Virtual World

    (b) BI

    (c) Software Integration

    (d) Web 2.0

6. The implementation of _____ saves the time and improves the operational efficiency.   [   ]

    (a) Software Integration

    (b) Social Network

    (c) Business Analytics

    (d) Business Intelligence

7. The process in which group of people in an organization collaborate on a specific data and discuss about the interpretation is known as _____.   [ ]

     (a)   Mobile Enterprise Network

     (b)   Collaborative Decision Making

     (c)   Business Intelligence

     (d)   RFID

8. _____ is a social networking which allows the users with common interest to collaborate and converse using mobile phones.   [ ]

     (a)   Mobile Social Networking

     (b)   Social Networking

     (c)   Mobile Community

     (d)   Mobile Enterprise Networks

9. _____ enables development of applications and interaction between the system components by employing tighter integration.   [ ]

     (a)   BI Implementation

     (b)   Web 2.0 Revolution

     (c)   Support Activity

     (d)   BI Software Integration

10. _____ allows the creation of avatars indicating their own identity which can help them in searching users of same interest.   [ ]

     (a)   Second Life

     (b)   RFID

     (c)   Reality Mining

     (d)   None of above

II.   Fill in the Blanks

1. Integration can be classified into two categories. They are _____ and _____.

2. The performance of _____ can be enhanced by providing various support activities.

3. _____ provides various light weight programming techniques.

4. _____ has helped many people in conducting business and earning profits.

5. _____ is the functional illustration of job contents, activities, duties and responsibilities of the job.

6. By using _____, an organization can recruit job seekers belonging to any part of the world.

7. The two RFID tags are _____ and _____.

8. _____ can be specified as a collection and examination of data resulted from machine sensed environment.

9. _____ is non routine activity that basically improvise discovery, innovation, learning and relations.

10. _____ software basically provides quantitative and qualitative analysis.

KEY

I. Multiple Choice

1. (b)
2. (c)
3. (c)
4. (b)
5. (a)
6. (d)
7. (b)
8. (a)
9. (d)
10. (a)

II. Fill in the Blanks

1. Functional Integration, Physical Integration
2. BI application
3. Web 2.0
4. Second Life
5. Job Description
6. e-recruitment
7. Active tag, Passive tag
8. Reality Mining
9. Brain storming
10. Social network analysis.

III.   Very Short Questions and Answers

Q1.   What is social networks?

Answer :

Social Networks are the networks that allow the users to interact with each other through messages, comments, images, posts etc.

Q2.   Define RFID.

Answer :

RFID stands for Radio Frequency Identification. RFID is an identification device used for describing several technologies.

Q3.   What is Facebook?

Answer :

Facebook is the most widely used social networking website having millions of users. It was found by Mark Zucker berg, Dustin Moskovitz and Chris Hugles at Harvard.

Q4.   What is myspace?

Answer :

In myspace every user has their own page in which they can create their profile with general information.

Q5.   What is Virtual World?

Answer :

Virtual world is a technology that creates an artificial environment that responds to, and is controlled by the behavior of the user.